AD_____

Award Number: DAMD17-99-1-9260

TITLE: The Importance of ATM Mutations and Polymorphisms in Breast Cancer and Radiation Sensitivity

PRINCIPAL INVESTIGATOR: Thomas A. Buchholz, M.D.

CONTRACTING ORGANIZATION: The University of Texas
MD Anderson Cancer Center
Houston, Texas 77030

REPORT DATE: October 2000

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2000	Annual Summary (1 Oct 99 - 30 Sep 00)

4. TITLE AND SUBTITLE
The Importance of ATM Mutations and Polymorphisms in Breast Cancer and Radiation Sensitivity

5. FUNDING NUMBERS
DAMD17-99-1-9260

6. AUTHOR(S)
Thomas A. Buchholz, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Texas M.D. Anderson Cancer Center
Houston, Texas 77030

E-Mail: tbuchhol@mdanderson.org

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

The primary objective of my 4-year career development award is to determine whether ATM heterozygosity is a significant contributor to both breast cancer risk and normal tissue injury resulting from radiation treatments of breast cancer. During the past year, we have begun full-length sequencing of ATM cDNA in these two patient cohorts. Our 3-year goal is to enroll 200 breast cancer patients and 50 breast cancer patients with injury. Over the first year, we have completed the sequencing of 89 breast cancer patients (21 with radiation injuries). Results of the sequencing studies have failed to find any protein truncating mutations that are known to be associated with the disease of ataxia telangiectasia. However, the sequencing has revealed a number of single base changes in the gene, the majority of which result in a change in an amino acid in the protein. It is not clear whether these changes affect the function of the protein or the stability of the mRNA. To evaluate this, we compiled a 960 individual control set of DNA's obtained from healthy volunteers. We have developed and validated allele specific oligonucleotide (ASO) assays to test the frequency of the polymorphisms in the patients and controls. The results of the ASO for the most prevalent polymorphisms are currently being analyzed.

14. SUBJECT TERMS

15. NUMBER OF PAGES
48

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

This research program investigates the relationship between an ATM mutation or polymorphism (ATM heterozygosity) and susceptibility to radiation-induced breast cancer and normal tissue injury following breast cancer radiation treatment. Over the 4-year period, we plan to study these potentially important relationships by sequencing ATM cDNA from 200 breast cancer patients and 50 breast cancer patients with a significant radiation injury. Our interest in studying ATM in breast cancer and radiation sensitivity first arose after obligate ATM heterozygotes (family members of children with ataxia telangiectasia) were noted to have a 5-fold greater risk of developing breast cancer than the general population (1). This led to estimates that 8% of all breast cancers occur in ATM heterozygotes (2). In addition, fibroblasts from obligate ATM heterzyogotes were noted to be radiosensitive compared to controls (3). Following cloning of the ATM gene, our preliminary work with cDNA sequencing and data from other investigators suggest that the protein-truncating mutations commonly found in patients in the disease with ataxia telangiectasia are uncommon in breast cancer patients. However, our data and that of others have identified a series of missense mutations that cause a single amino acid change in the protein product of the ATM gene. This has led some authors to propose that the changes in the ATM gene may be different for breast cancer development and the disease of ataxia telangiectasia. Specifically, they suggest that it is single nucleotide changes in the gene that may have an association with breast cancer development (4). However, it is yet unclear whether these mutations result in functional consequences that may predispose to breast cancer or whether they represent inconsequential polymorphisms. In the upcoming years of this program we will be testing whether these polymorphisms are more frequently found in breast cancer patients compared to controls. We have developed allele-specific oligonucleotide assays to compare the frequency of specific single base nucleotide changes seen in our sequenced breast cancer population to the frequency in ethnically-matched controls drawn from a group of 960 individuals who donated blood for this project during community-based blood drives. In addition, we plan on developing *in vitro* assays to investigate whether missense mutations/polymorphisms we have identified result in functional consequence. The phenotypes we plan to evaluate are cellular radiosensitivity and capacity for DNA damage repair.

Body

The first specific aim in our statement of work was to establish the incidence of ATM heterozygosity in 200 breast cancer patients. This aim was to be completed over the first 3 years of the funding period. To achieve this aim, we had developed a cDNA sequencing process for the ATM gene prior to the time of awarding of this grant. The sequencing process was verified by confirming the presence of a mutation in two obligate ATM heterozygotes (parents of an individual with the disease of ataxia telangiectasia). During the first year of the funding period, task 1 of specific aim 1 was to begin sequencing the cDNA of breast cancer patients. Over this first year, we enrolled 101 breast cancer patients onto our institutional protocols studying the ATM gene. As specified in this task, the coding of all clinical and epidemiological information was abstracted and recorded for all of these patients. From these samples, RNA was isolated and reverse transcribed into cDNA and then sequenced for mutations in ATM. Sufficient information from sequencing of ATM cDNA was obtained in 89 subjects.

Our first specific aim focused on studying a nonselected breast cancer patient population. Our second specific aim focused on a second set of breast cancer patients selected because they had experienced a significant normal tissue radiation injury. Of the 89 subjects sequenced thus far, 68 breast cancer patients were in the nonselected population. No protein truncating mutations were found in this cohort. A total of 30 patients had variation of the sequenced gene compared to the GeneBank sequence. Specifically, a total of 41 single base changes were detected in the 30 patients. Nine patients had 2 or more single nucleotide changes. From these patients, we identified 4 missense mutations/polymorphisms that occurred repeatedly in 3 or more of the 89 patients. The specific polymorphisms and their frequency are listed below:

1. Below is a portion of the ATM cDNA sequence with a C to G single nucleotide change:

```
  1 tttttagtag agacagggtt tcaccatgtt ggtcaggctg gtcgaactcc tgaccttgtg
 61 atcctcccac cttggcctcc caaagtgctg ggattacagg cgtgagccac tgcgtccagt
121 gtaaattata cttttatttt aatcctgcta ctactgcaag caaggcaaac atttttgtgt
181 tacagcatta cttgtataga ttttaagaaa atctcatttt aaatacggaa atgttaagaa
241 aaattattgt gcctttgacc agaatgtgcc tctaattgta cagttaaatc taactataaa
301 tactgcagta taaaataatt atatacacat tttttcacac ctctttctct ctatatatgc
361 atatatacat atacatatat atacctatat gtatttttt tacagacagt gatgtgtgtt
421 ctgaaattgt gaaccatgag tctagtactt aatgatctgc ttatctgctg ccgtcaacta
481 gaacatgata gagctacaga acgaaaggta gtaaattact taaattcaat ttttnccttg
541 aaatgtgtga ttagtaaccc attattattt tccttttttat tttcagaaag aagttgagaa
601 atttaagcgc ctgattcgag atcctgaaac aattaaacat ctagatcggc attcagatt(c/g)
661 caaacaagga aaatatttga attgggatgc tgtttttagg tattctattc aaatttattt
721 tactgtcttt atttttctct ttcatattta tttctgttgt gatattactt ttgtgtgtaa
781 gtcttaacat ttatctttgc ttcctatata tcattatgcc ttgcatatga atttggcatt
841 taatatttat ccaaaacata atttttaaag gttgttcata tagaaactta aaaattataa
901 attatttctt ccaataaaat gtttt
```

The resulting change in the protein is - Ser49Cys

This change was seen in 4 of 56 patients (the denominator does not total the total patient number because the entire gene sequence was not obtained in every patient sample).

2. Below is a portion of the ATM cDNA sequence with a G-A single nucleotide change:

```
  1 aacttttgat acctttttcc ctcttctatc atatcagaaa actccttctg ttactaacgc
```

```
 61 ttttcaactc tgaaattgta tttaaagtgg tgatataatt tcatttgtaa gcaatattct
121 gtttcattat ggtaatggcc tagactggaa ataaacagtt acagtgtcac taacatatat
181 atttgatatt gatatactag cctagtgtgg ttttttaaac accacctaat acatgttttt
241 tgtttgtttt tttagcagta tgttgagttt atggcagatt aatctatcat cttttagaaa
301 tttaatatgt caacggggca tgaaaatttt aagtaaaatg tattaatttt actcattttt
361 actcaaacta ttgggtggat ttgtttgtat attctaggtg aaaactgact tttgtcagac
421 tgtacttcca tacttgattc atgatatttt actccaa(g/a)at acaaatgaat catggagaaa
481 tctgctttct acacatgttc agggattttt caccagctgt cttcgacact tctcgcaaac
541 gagccgatcc acaacccctg caaacttgga ttcaggtatt ctattaaatt tttaacatta
601 atactgtaaa ctcagttcta gagaaagatg gatttaagat ggaatcccac taaaagcact
661 ttacaggatt aaatctataa cctctaaatt tgtttcttca tctatggaat ggagataaaa
721 gttgccaaca gttgcaacaa gtttccaatg aaataatgtg tgtaaagtgc ctaggatagt
781 acttgatgta tagtattccc tcagcacatt tcggctattg ataatgggtc aactaattga
841 gctttcaata tgtgtcaggc actgtgcttg cactggcaat attaatgtga aaaagaaaca
901 cccacattct agcaatggaa aaaacaa
```

The resulting change in the protein is - Asp1853Asn

This change was seen in 15 of 59 patients.

3. Below is a portion of the ATM cDNA sequence with a T to C single nucleotide change:

```
  1 ctcaagaggg caaaaggatg cgtaactgct tacctatacc agatgttaaa ggtttttaaa
 61 cctatgctct ttacttcctc tgcttggtga aaaaagggat atgtttgcag acaatgttat
121 gcttaacatt tatatctggt gtttttaaaa atactttctg aatttgcctt tgagattata
181 acttgtattt tttctctatc tattagtaaa atttgctact gaataatgac atttgatata
241 agtaggtctc aaagtccgaa gaagagaagc atttaaaaga ataatctatt aattatataa
301 gtagtctttg aatgatgtag atactaggtt aatgttttcc tttgtaatat attgctaata
361 catataaggc aaagcattag gtacttggtt tatatattaa agatcttact ttcttgaagt
421 gaacaccacc aaaaagataa agaagaactt tcattctcag aagtagaaga actatttctt
481 cagacaactt ttgacaagat ggacttttta accattgtga gagaatgtgg tatagaaaag
541 caccagtcca gtattggctt ctctgtccac cagaatctca aggaatcact ggatcgctgt
601 cttctgggat tatcagaaca gcttctgaat aattactca(t/c) ctgaggtgag atttttaaa
661 aaaagaacta agcttatata tgattcaact ttggtaaact gttaggaagg agaaataggg
721 gcaggaaaaa cagcaaggat ggtgggaggc ttcattttaa aagcaaagtg gcagtaaagg
781 gctctaaatt ggacaactta gcataattaa aggaaaactc aagaataata atttgagtac
841 ttcctttgta ctggaaatta tggtagacat aaaataattc cttgtgtagg ttagtgagga
901 atagtaagag tttgagcata gggattatat gatgaaaaaa acctctaaat acaaaggagg
961 gaaatgttac agtaatagaa aagaacacga tgtaaacaaa tctaatagat tttgg
```

The resulting change in the protein is – Ser707Pro

This change was seen in 4 of 57 patients.

4. Below is a portion of the ATM cDNA sequence with the detected C to G single nucleotide change:

```
  1 ctcccaaagt gttgggatta caggtgtgag ccagtgcacc cagcctgaat gtggttttga
 61 aatcttcaat ataccaaaca aagaaatctt caatatacca aacaaagaaa tttcttttaa
121 agtaagttta ctgtaatgta gtttagccat tgtatggtag cccccaaaaa aggacataat
181 ggtataaaac aaataatata cgctaaaatg aattctttta cactaatttc ttttagcttg
241 aatttttggc aaggtgagta tgttggcata ttccacataa tgacaaataa gtttagcaca
301 gaaagacata ttggaagtaa cttataataa cctttcagtg agttttctga gtgcttttat
```

```
361 cagaatgatt atttaacttt ggaaaactta cttgatttgc aggcatctaa caaaggagag
421 gaaatatata ttctctgtaa gaatggccct agtaaattgc cttaaaactt tgcttgaggt
481 gagtttttgc atttttttag taagatctcc attgaaaatt ttaaagcagt ctttgtttgt
541 taatgagtaa tttttctcta tttcatattt aaccacagtt cttttcccgt aggctgatc(c/g)
601 ttattcaaaa tgggccattc ttaatgtaat gggaaaagac tttcctgtaa atgaagtatt
661 tacacaattt cttgctgaca atcatcacca agttcgcatg ttggctgcag agtcaatcaa
721 taggtaatgg gtcaaatatt catgaagtat ttggaatgct gcagatggca gtagaatgtc
781 ttacatagta acagctcaca gttgcaatat taaaaatagc taacacttgt tgagtatata
841 cggtgtgcct ggcatttatg tttattctta attcttatac ttctgtcact tagattctat
901 tatttccttc aatttataaa tga
```

The resulting change in the protein would be – Pro1054Arg

This change was seen in 1/61 patients.

Our second specific aim in our statement of work was to establish over the first 3 years of the funding period the incidence of ATM heterozygosity in 50 breast cancer patients experiencing a significant acute or late normal tissue radiation injury. In the first year of this award, we have sequenced the cDNA of 21 breast cancer patients who had radiation complications. No protein truncating mutations were found in this cohort. A total of 5 patients had sequence variation of ATM cDNA compared to that listed in GeneBank. Specifically, a total of 8 single base changes were detected in the 6 patients. Two patients had 2 single nucleotide changes. The frequency of these polymorphisms / missense mutations were:

1. Ser49Cys (shown above): 1/16
2. Asp1853Asn (shown above): 5/16
3. Ser707Pro (shown above): 0/16
4. Pro1054Arg (shown above): 2/19

The third specific aim outlined in our original statement of work was to compare the frequency of identified ATM polymorphisms / mutations in the patient samples to the frequency of these mutations in an ethnically matched control set. Ethnically matched controls were to be obtained from a general population of individuals who did not have a personal history of cancer. During year 1 of the award, our goal was to establish a control bank of 1,000 DNA samples from individuals. Working with our colleagues from the Department of Epidemiology, we have completed this task and have isolated DNA from a control sample of 960 individuals. Our second task during this period was to develop ASO assays for rapidly screening the DNA bank for specific ATM polymorphisms. This work is currently complete for three of the four missense mutations above (numbers 1, 2, and 4). The comparative frequency testing for these three polymorphisms has been performed. Over the remaining funding period, we plan to continue developing the ASO assay for the third polymorphism above as well as additional polymorphisms that are repetitively found in the patient samples. The final task of specific aim 3 is to compare the frequency of ATM polymorphisms in the patient samples with case controls from the stored DNA bank. For this task, we have developed a new collaboration with Ranaji Chakarabarty, Ph.D., Professor in the Human Genetics Center at the University of Texas Houston School of Public Health. Doctor Chakarabarty has specific expertise in population genetics and will serve as an important mentor in the data analysis. Based on this strategy, the appropriate sample size to validate any unique polymorphism can be determined and a more efficient and rapid screening of a large breast cancer population using the ASO assay rather than complete gene sequencing can be accomplished.

In addition to the specific tasks outlined below, we have developed additional collaborations focused on ATM research that directly parallels the specific aims of this study. Specifically, in a collaborative effort with Ms. Penelop Bonnen, Ph.D., and David Nelson, Ph.D. of the Department of Molecular and Human Genetics of Baylor College of Medicine, we established haplotypes in the ATM gene. This work was important in establishing the feasibility of using haplotype association studies to detect individuals with a genetic predisposition for a disease. A manuscript of this work is included in Appendix A and will be published in Am J Hum Gen (67:,2000). The support from the Career Development Award was acknowledged in the manuscript.

In addition to this collaboration, the first year funding from the Career Development Award has helped to further development of my career in breast cancer radiation oncology. Some of the insights gained from our research inspired me to further explore the relationship between breast cancer development and cellular radiosensitivity. In separate study, we demonstrated that lymphocytes from patients with bilateral breast cancer had a greater number of radiation-induced chromatid breaks than controls. The initial report from this work resulted in a peer-reviewed oral presentation at the First International Conference on Translational Research and Pre-Clinical Strategies in Radio-Oncology, Lugano, Switzerland (March, 2000) and a manuscript accepted for publication in the Int J Radiat Oncol Biol Phys. The support from my DOD career development award was also acknowledged in this manuscript. This manuscript is included in Appendix B.

Key research accomplishments (Year 1)

- Sequenced ATM cDNA from 89 breast cancer patients
- Identified 4 single nucleotide base changes resulting in an amino acid change in the protein present in multiple patients
- Established a control bank of DNA from 960 individuals without a cancer history for population comparison studies
- Developed an allele specific oligonucleotide assay for 3 single nucleotide base changes in the ATM gene
- Established a collaborative effort that demonstrated the feasibility of using haplotype association studies with the ATM gene.
- Demonstrated that bilateral breast cancer patients have an increased number of radiation-induced chromatid breaks compared to controls showing the feasibility of a phenotype rather than genotype assay to predict breast cancer risk.

Reportable Outcomes

- manuscripts, abstracts, presentations:

1. Buchholz TA, Wu XF: Radiation-induced chromatid breaks as a predictor of breast cancer risk. . Int J Radiat Oncol Biol Phys, 2000.

2. Bonnen PE, Story MD, Ashorn CL, Buchholz TA, Wiel MM, Nelson DL: Haplotypes at ATM identify coding sequence variation and indicate a region of reduced recombination. Am J Hum Gen, 2000.

3. Buchholz TA, Wu XF: Radiation-induced genomic instability as a predictor for the risk of breast cancer development. Int J Rad Oncol Biol Phys 46(3):766 (#198), 2000.

– patents and licenses applied for and/or issued: none

– degrees obtained that are supported by this award: none

– development of cell lines, tissue or serum repositories

1. established a control bank of DNA from 960 individuals with no cancer history to serve for population frequency testing

– informatics such as databases and animal models, etc:

1. established a database for enrolled patients and controls that cover patient demographics, cancer history, and toxicity from radiation treatment.

– funding applied for based on work supported by this award: none

– employment or research opportunities applied for and/or received on experiences/training supported by this award: none.

Conclusion

Over the first year of this 4-year program, we have been successful in sequencing ATM cDNA in breast cancer patients and breast cancer patients with radiation injury. Based on these studies, we conclude that these populations have multiple single nucleotide base changes that may have contributed to their disease and treatment-related toxicity. To explore this relationship further, we have successfully developed allele specific oligonucleotide assays to test population frequency of single nucleotide changes in patients and ethnically-matched controls.

In addition to continuing the planned studies outlined in our original statement of work and body of this report, we hope to determine some of the single base changes result in cellular radiosensitivity and deficiency in DNA damage repair. We have obtained institutional review board approval of a protocol to obtain skin fibroblasts from individuals in whom our sequencing studies have identified single nucleotide changes. The fibroblasts will be grown in culture for in vitro assays of radiosensitivity, DNA damage repair, and biochemistry assays.

Establishing a relationship between ATM and breast cancer development and/or normal tissue toxicity following breast cancer radiation treatment would be a significant contribution to breast cancer research. If our data identifies an association between specific single nucleotide changes and breast cancer, further studies will be needed to determine the excess risk. Ultimately, allele specific oligonucleotide assays could serve as a simple tool to screen large populations of women to further define their breast cancer risk and risk of treatment-related toxicity.

References

1. Swift M, Peitnauer P, Morrell D, Chase C: Breast and other cancers in families with ataxia telangiectasia. New Engl J Med 316:1289-1294, 1987.

2. Swift M, Morrell D, Massey R, Chase C: Incidence of cancers in 161 families affected by ataxia telangiectasia. New Engl J Med 325:1831-1836, 1991.

3. Kinsella TJ, Mitchell JB, McPherson S, Russo A, Tietze F: In vitro X-ray sensitivity in ataxia-telangiectasia homozygote and heterozygote skin fibroblasts under oxic and hypoxic conditions. Cancer Res 42:3950-3956, 1982.

4. Gatti RA, Tward A, Concannon P: Cancer risk in ATM heterozygotes: a model of phenotypic and mechanistic differences between missense and truncating mutations. Mol Genet Metab 68(4):419-23, 1999.

Am. J. Hum. Genet. 67:000–000, 2000

Haplotypes at *ATM* Identify Coding-Sequence Variation and Indicate a Region of Extensive Linkage Disequilibrium

Penelope E. Bonnen,[1] Michael D. Story,[2] Cheryl L. Ashorn,[2] Thomas A. Buchholz,[3] Michael M. Weil,[2] and David L. Nelson[1]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, and Departments of [2]Experimental Radiation Oncology and [3]Radiation Oncology, MD Anderson Cancer Center, Houston

Genetic variation in the human population may lead to functional variants of genes that contribute to risk for common chronic diseases such as cancer. In an effort to detect such possible predisposing variants, we constructed haplotypes for a candidate gene and tested their efficacy in association studies. We developed haplotypes consisting of 14 biallelic neutral-sequence variants that span 142 kb of the *ATM* locus. *ATM* is the gene responsible for the autosomal recessive disease ataxia-telangiectasia (AT). These *ATM* noncoding single-nucleotide polymorphisms (SNPs) were genotyped in nine CEPH families (89 individuals) and in 260 DNA samples from four different ethnic origins. Analysis of these data with an expectation-maximization algorithm revealed 22 haplotypes at this locus, with three major haplotypes having frequencies ≥.10. Tests for recombination and linkage disequilibrium (LD) show reduced recombination and extensive LD at the *ATM* locus, in all four ethnic groups studied. The most striking example was found in the study population of European ancestry, in which no evidence for recombination could be discerned. The potential of *ATM* haplotypes for detection of genetic variants through association studies was tested by analysis of 84 individuals carrying one of three *ATM* coding SNPs. Each coding SNP was detected by association with an *ATM* haplotype. We demonstrate that association studies with haplotypes for candidate genes have significant potential for the detection of genetic backgrounds that contribute to disease.

Introduction

Qualifying and quantifying the genetic contribution to the etiology of common complex disease remains one of the great quests of modern medical genetics. The complexity of multifactorial diseases challenges the paradigms and tools of conventional genetic research. Traditional methods of genetic analysis do not have the statistical power or sensitivity for the task of teasing out a genetic contribution when it is subtle or when several genes may be working together (Risch and Merikangas 1996). Genomewide association studies, as well as population studies with candidate genes, have been touted as possible alternatives to linkage analysis (Risch and Merikangas 1996; Collins et al. 1997; Kruglyak 1999; Risch 2000). These approaches focus on finding either a causative variant or a genetic variant closely linked with the disease phenotype. Some studies utilizing single-nucleotide polymorphisms (SNPs) have succeeded in detecting the risk for disease, notably in the case of the apolipoprotein type E (apoE) gene and both coronary artery disease (Boerwinkle et al. 1996) and Alzheimer disease (Strittmatter and Roses 1995). These studies were able to directly assess the risk conferred by known apoE functional variants. In some other cases, however, the attempt to correlate single-locus alleles with phenotypes have produced mixed results (Josefsson et al. 1998; Kraft et al. 1998; Storey et al. 1998).

Haplotype association with disease by the linkage disequilibrium (LD) approach has been used successfully for the identification of genomic regions containing loci responsible for disease phenotypes (MacDonald et al. 1992; Yu et al. 1996). The same principle can be applied by use of haplotypes of biallelic markers to detect disease association. Using several SNPs distributed across 100–200 kb should result in statistical sensitivity that is greater than that in studies using fewer loci. Another strength of such an approach is the ability to use purely epidemiological populations for detection of chromosomal backgrounds lending risk for disease.

All of these approaches are, to one extent or another, dependent on LD. An understanding of LD relationships between markers will inform the efficacy and design of future LD-based strategies for detection of genetic contributions to common disease. Simulation studies have estimated the length of useful LD to be as low as 3 kb (Kruglyak 1999). Recent investigations support the no-

Received August 22, 2000; accepted for publication September 21, 2000; electronically published November 14, 2000.

Address for correspondence and reprints: Dr. David L. Nelson, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. E-mail: nelson@bcm.tmc.edu

0002-9297/2000/6706-0010$02.00

tion that LD varies throughout the genome (Collins et al. 1999; Taillon-Miller et al. 2000) and that it can extend to considerable lengths, such as several hundred kilobases (Collins et al. 1999; Eaves et al. 2000; Moffatt et al. 2000; Taillon-Miller et al. 2000). Reports of such extreme differences indicate the need for further study of the extent and nature of LD.

Allelic variation leading to functional variants of genes may predispose to risk for seemingly sporadic cases of common disease (Lander 1996; Collins et al. 1997). Here we describe a strategy for exploring the possible effects of functional variants of genes involved in familial cancers. We use a resequencing approach to detect SNPs across a large (184 kb) genomic region containing the *ATM* gene. *ATM* is responsible for the autosomal recessive disease ataxia-telangiectasia (A-T) (Savitsky et al. 1995). A-T is characterized by cerebellar ataxia, oculocutaneous telangiectasia, immune deficiency, sensitivity to ionizing radiation, increased incidence of tumors, and chromosomal instability (Gatti et al. 1991). A-T heterozygotes may be at increased risk for development of cancers, most prominently—and controversially—breast cancer (Swift et al. 1987, 1991; Morrell et al. 1990; Stankovic et al. 1998; Gatti et al. 1999). With carrier frequencies estimated to be from 0.5% to >1% (Swift et al. 1986; Gatti et al. 1999), assessment of cancer risk for this population is a compelling endeavor. In addition, the *ATM*-gene product is centrally involved in cellular responses to DNA damage, including DNA double-strand break repair and signaling leading to cell-cycle arrest and apoptosis (reviewed in Rotman and Shiloh 1999). We genotyped 295 individuals from four ethnic groups, for 14 SNP markers that spanned 142 kb. An expectation-maximization algorithm estimated 22 *ATM* haplotypes from these data. Tests for recombination and LD revealed (*a*) no evidence for recombination in the white European American study population and (*b*) perfect disequilibrium extending the full length marked by these SNPs. We then conducted a model association study with these haplotypes and a population of samples that possessed one of three different coding SNPs (cSNPs) in the *ATM* gene. The results of this study provide strong support for the utility of complex SNP haplotypes as a means to detect polymorphisms in a population-based sample.

Subjects and Methods

Human Subjects

For SNP discovery, genomic DNA from five unrelated white European Americans was sequenced. This DNA was extracted from lymphoblast and fibroblast cell lines. For SNP genotyping, individuals from four ethnic groups were sampled: African American ($n = 71$), Asian American ($n = 39$), white European American ($n = 77$), and Hispanic ($n = 73$). All ethnic samples (self-described ethnicity) were part of a collection of 941 DNA purified samples from anonymous blood donors in community-based blood drives in southeastern and central Texas. Samples analyzed in the model association study were also from this DNA collection. Members of nine CEPH families were also analyzed. In all families, four grandparents, two parents, and four children were examined; since two of these families share a grandparent, 89 individuals were genotyped, and the number of segregating chromosomes is 70.

Samples from Great Apes

Six great-ape samples were genotyped in this study: two from common chimpanzees (*Pan troglodyte*), one from a bonobo (*P. paniscus*), two from western lowland gorillas (*Gorilla gorilla*), and one from an eastern lowland gorilla (*G. g. graueri*).

PCR and Sequencing Primers

Primers for DNA amplification and sequencing were designed by MacVector, version 6.0.1. The 184-kb genomic sequence of *ATM* was masked for repetitive sequence, by Repeat Masker. Thirty-six primer sets were designed to amplify regions containing little or no repeat sequence, distributed evenly throughout the sequence. Primers were selected that met strict criteria for melting temperature and that amplified regions containing very little or no repeat sequence. The same primers were used for PCR and sequencing reactions and are listed in Appendix A.

PCR Amplification of Genomic DNA

Genomic DNA from five unrelated individuals was amplified by means of 29 of the 36 primer sets mentioned above. The 50-μl reactions included DNA (200 ng), standard PCR buffer, dNTPs (0.1 mM each), *Taq* (0.5 μl; Perkin-Elmer), and primers (1 μM each). PCR was performed in a Perkin Elmer 9700 analyzer, with an initial denaturation at 95°C for 5 min, followed by 30 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s, and a final step at 72°C for 7 min. For all amplicons, 6 μl of PCR product was run on a 1.5% agarose gel.

DNA Sequencing

PCR products were purified and sequenced. Preparation of DNA for sequencing included incubation of ~60 ng of PCR product with shrimp alkaline phospha-

tase (2 U; Amersham) and exonuclease I (10 U; Amersham) at 37°C for 15 min, followed by enzymatic inactivation at 80°C for 15 min. Sequencing of each PCR product was performed with the Thermo Sequenase [^{33}P]-radiolabeled terminator-cycle sequencing kit (Amersham Pharmacia), according to the manufacturer's instructions. Sequencing reactions were performed in a Perkin Elmer 9700 analyzer, with an initial denaturation at 95°C for 1 min, followed by 35 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min. Samples were run on 6% polyacrylamide gels, fixed for 15 min in 5% acetic acid/20% methanol, and dried.

Multiplex PCR

Sequencing revealed 17 SNPs in 15 different regions of the gene. These 15 PCR amplicons were multiplexed into two PCR reactions. Multiplex group 8 amplifies eight fragments, and Multiplex group 7 amplifies seven fragments. The 50-μl reactions for group 7 included DNA (400 ng), standard PCR buffer (2 ×), dNTPs (0.2 mM each), and *Taq* (0.5 μl; Perkin-Elmer). The 50-μl reactions for group 8 included DNA (400 ng), standard PCR buffer (1.8 ×), dNTPs (0.2 mM each), and *Taq* (0.5 μl; Perkin-Elmer). Primers include some of those originally designed for sequencing and some of those newly designed to alter the size of the amplicons. Products were separated by ≥20 bp, so that they could be resolved from one another on a 2.5% agarose gel. Multiplex PCRs were checked to have amplified all products, by running 6 μl of product on a 2.5% agarose gel. The concentrations and primer sequences used for PCR are listed in Appendix B.

Allele-Specific Oligonucleotide (ASO) Hybridizations

Genotypes for each SNP were determined in all sample populations, by ASO hybridizations. ASO hybridizations were performed as described by DeMarchi et al. (1994). We performed ASO hybridization for 14 SNPs for each individual typed. These 14 SNPs were chosen from the original 17 because they perform consistently well under standard ASO-hybridization conditions. Hybridizations were performed under conditions that allowed for annealing of only the probe that is an exact match for the substrate DNA. Genotypes for SNPs were read on at least two independent occasions. The sequences of the ASO-hybridization probes are listed in Appendix C.

Estimation of Haplotypes and Frequencies

Haplotypes and their frequencies were estimated on the basis of unphased genotype data, by the computer program EMHAPFRE. Described in the work of Excoffier and Slatkin (1995), EMHAPFRE uses an expectation-maximization algorithm that determines the maximum-likelihood frequencies of multilocus haplotypes in diploid populations. Only individuals who were scored for all 14 SNPs were included in the data analysis.

Haplotype Assignment to Genotype Data

A short script written in Microsoft Excel Visual Basic and named "Assign" was used to assign genotypes to individual samples. The script was given, as input, the list of haplotypes produced by EMHAPFRE and the raw unphased genotype data. It produces a list of samples input, with a pair(s) of haplotypes that satisfies the genotype data assigned to each; in cases in which multiple pairs of haplotypes were listed, one pair is chosen, by use of a haplotype frequency–based method. A probability is calculated for each haplotype pair, by multiplication of the haplotypes' frequencies in the control population. The haplotype pair with the highest probability is assigned to the individual.

Statistical Analysis for Recombination and LD

To test for recombination, we used the four-gamete test and the Hudson and Kaplan (1985) recombination statistic, R. For a given haplotype AB, mutation may result in either Ab or aB. Haplotype ab arises only in the case of either recombination or repeat mutation. The four-gamete test was executed on unphased genotype data, in a pairwise fashion, across all SNP loci. On the basis of the resulting matrix of the four-gamete test, R estimates the location and number of recombination events that have occurred in the sample.

Initial LD analysis was computed by performance of pairwise comparisons for all SNP loci. Fisher's exact test was used to determine significance levels. SNPs having a minor-allele frequency of .05 were excluded from LD analyses. LD statistic D is a pairwise comparison of gametic frequencies such that D = p11p22–p12p21. D', the relative disequilibrium, is $D' = D/|D|\text{max}$, where $|D|\text{max} = \max(\text{p1p2,q1q2})$ if $D < 0$ and $|D|\text{max} = \min(\text{q1p2,p1q2})$ if $D > 0$. D' ranges from 1 to −1, and this range is not influenced by allele frequency.

All recombination and LD statistics were generated by the software program DnaSP 3.00 (written by J. Rozas and R. Rozas, University of Barcelona).

Statistical Analysis for Association Study

Testing for significance in the model association study was done by use of contingency tables for independence. P values for significance of association at the haplotype level were determined by use of 2 × 2 tables and 3 × 3 tables for the genotype level. Significance values refer to a one-sided test.

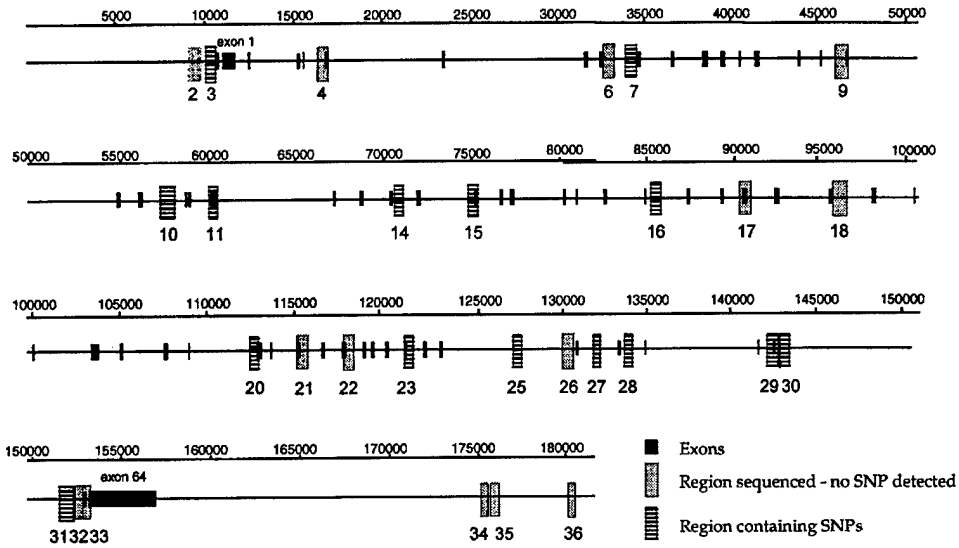


Figure 1 Schematic of *ATM*. The 184 kb of the *ATM* locus is illustrated, with the 64 exons represented by black boxes. Twenty-nine ~500-bp regions were amplified by PCR in five unrelated individuals. These regions were sequenced and found to contain 17 SNPs.

Results

SNP Discovery

Our initial objective was to discover common neutral sequence variants spanning the length of the *ATM* gene. A gel-based resequencing strategy was employed to detect SNPs at the *ATM* locus. Genomic DNA of five unrelated individuals was amplified, by PCR, for [^{33}P]-radiolabeled sequencing. For detection of markers spanning the entire locus, PCR primers were designed for amplicons dispersed approximately evenly throughout the 184-kb genomic region containing the gene (fig. 1). Approximately 13.5 kb of the 184-kb total sequence was read in each individual. The nucleotide diversity, π, calculated for this sequence data was .00057. Seventeen SNPs were found, which span 142 kb and all of which are located in introns (table 1). This yielded an average of 1 SNP/794 nucleotides sequenced.

Genotyping and Haplotype Development

To begin construction of haplotypes from these SNPs, we genotyped nine three-generation CEPH families (Dausset et al. 1990). By using three-generation families, we could determine haplotypes from genotype data, through inference. This allowed us both to determine the efficacy of the computer algorithm used to predict haplotypes (see below) and to optimize our genotyping assay. We performed ASO hybridization on nine CEPH families (89 individuals; 70 chromosomes), for 14 of the original 17 SNPs. These 14 SNPs were chosen from the original 17 because they performed consistently well under standard ASO-hybridization conditions.

We then used two different methods for deciphering the haplotypes derived from the genotype data, in a side-by-side comparison. First, haplotypes were inferred by

Table 1

Seventeen *ATM* Noncoding SNPs Detected by Resequencing

SNP[a]	Location in Genomic Sequence with GenBank Accession Number U82828
Prior to 5′UTR t→a[b]	10182
IVS8−356t→c	34293
IVS19−1276a→g	57469
IVS21−77t→c	60136
IVS26+491c→g[c]	71049
IVS27−193c→t[c]	75083
IVS34+754g→a	85811
IVS46−257a→c	112721
IVS55+186c→t	121819
IVS57+3570t→c	127195
IVS58+997g→a	132032
IVS59+414g→t[c]	133986
IVS61−55t→c	142611
IVS62+60g→a	142789
IVS62+424g→a	143153
IVS62−973a→c	151964
IVS62−694c→a	152243

[a] Nomenclature is according to the guidelines recorded by the Ad Hoc Committee on Mutation Nomenclature (1996).

[b] This SNP is named in reference to the genomic sequence having GenBank accession number U82828 because of the highly variable nature of the 5′UTR.

[c] Not used in genotyping or haplotype analysis.

Table 2

***ATM* Haplotypes of 295 Humans from Five Ethnic Groups and of Three Species of Great Apes**

		Frequency in Humans[a]					
Haplotype	Sequence	Overall (n = 295)	African American (n = 71)	Asian American (n = 39)	White European American (n = 77)	Hispanic American (n = 73)	CEPH (n = 35)
1	ACTCTACTTCCCTC	.002				.007	
2[b]	ACTCTACTTCTTTC	.313	.190	.500	.292	.315	.394
3	ACTCTCCTTCTTTC	.037			.065	.048	.061
4	ACTTCACTCCTCTC	.002	.007				
5	ACTTTACTCTCCTC[c]	.002					
6[b]	ACTTTACTTCCCTC	.066	.218	.013	.013	.027	.015
7	ACTTTACTTCTTTC	.019	.077				
8	ATTCTACTTCTTTC	.012	.007	.051		.007	
9	ATTCTCCTTCTTTC	.000		.013			
10	ATTTCACTCCCCTC	.002	.007				
11	ATTTCATCCTCCCC	.002		.013			
12	TCTCTACTTCTTTC	.007				.021	.015
13	TCTTCACTCTCCTC	.010	.035			.007	
14	TCTTCATCCTCCCC	.002				.007	
15[b]	TTCTCACTCTCCTA	.090	.028	.013	.175	.041	.227
16	TTTCTATCCTCCCC	.005		.017		.007	
17[b]	TTTTCACCCTCCTC	.100	.141	.068	.097	.110	.015
18	TTTTCACCCTCTTC	.002				.007	
19	TTTTCACTCCTTTC	.002	.007				
20	TTTTCACTCTCCTA	.002				.007	
21[b]	TTTTCACTCTCCTC	.048	.162	.013	.006	.027	
22[b]	TTTTCATCCTCCCC	.277	.113	.291	.351	.363	.273
	TTTCTACCCTCCTC[c]	...		.009			
	ACTTTACCCTCCTC[c]	...	.007				
Total		1.000	1.000	1.000	1.000	1.000	1.000
			Frequency in Great Apes[d]				
			Chimpanzee (n = 2)	Bonobo (n = 1)	Gorilla (n = 3)		
1	TCTTTACTCTCCTC		.750	1.000	.000		
2	TCTTTACTCTCTTC		.250	.000	.000		
3	TATTTACTCTCCTC		.000	.000	1.000		

[a] Samples were genotyped by ASO hybridization, then haplotypes and their frequencies were estimated from unphased genotype data, by the EM algorithm EMHAPFRE.

[b] Haplotype present in all four ethnic groups studied.

[c] Low-frequency haplotypes in which some differences were seen in the combined data set and in individual ethnic populations.

[d] Samples were genotyped by ASO hybridization and fluorescent sequencing.

hand. We began with homozygotes and predicted other haplotypes on the basis of transmission and by establishing the phase through the pedigrees. Seven haplotypes were identified in the sample of CEPH families. Subsequently, we subjected the same data set to an expectation-maximization algorithm, to estimate haplotypes and their frequencies. The computer program EMHAPFRE is a maximum-likelihood program developed to predict multilocus haplotypes from unphased genotype data (Excoffier and Slatkin 1995). It produces both a list of haplotypes and their estimated frequencies in the input sample population. The haplotype predictions from EMHAPFRE were in complete accordance with those that had been inferred manually, giving us confidence that this program was suitable for data of this nature.

Haplotype and Allele Frequencies

To determine frequencies of haplotypes and of individual SNPs in different ethnic populations, we performed ASO hybridization on anonymous African American (n = 71), Asian American (n = 39), white European American (n = 77), and Hispanic (n = 73) DNA samples collected in central and southeastern Texas. Genotype data were analyzed by the EMHAPFRE program. For the total population, 22 haplotypes and their frequencies were predicted by EMHAPFRE (table

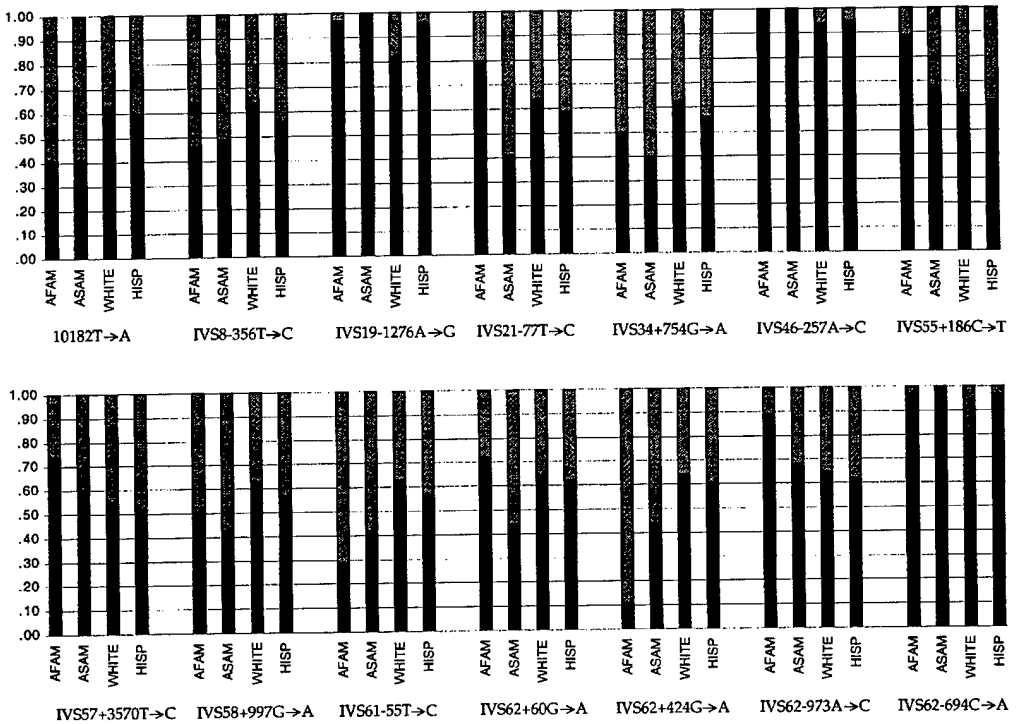


Figure 2 ATM SNP allele frequencies for 14 ATM SNPs in each of four ethnic groups. A total of 260 individuals (71 African American, 39 Asian American, 77 white European American, and 73 Hispanic) were genotyped by ASO hybridization.

2). Three predominant haplotypes were found at frequencies ≥10%. An independent study that examined neutral sequence variants at the *ATM* locus also found three major haplotypes (Li et al. 1999).

The majority of SNPs identified in this study have a frequency, in all ethnic groups, of ≥25% (fig. 2). Of the 14 SNPs, 3 (IVS19−1276a→g, IVS46−257a→c, and IVS62−694c→a) have a minor-allele frequency of <10% in most ethnic groups. SNP frequencies vary across ethnic groups. Three SNPs (IVS55+186c→t, IVS62+424g→a, and IVS62−973a→c) have a frequency of 11% in African Americans while being present at a frequency of >30% in all other ethnic groups. SNP IVS46−257a→c was not found in the samples from African Americans. Of the three low-frequency SNPs, two (IVS19−1276a→g and IVS62−694c→a) have a frequency of >18% in the white European American population and of <6% in the others. This is not surprising, given that the original five samples used for SNP detection were white European Americans.

To begin to describe the haplotype phylogeny at the *ATM* locus, we wanted to determine what haplotypes were present in each ethnic population. The genotype data were analyzed, by EMHAPFRE, as four separate data sets segregated by ethnic group. However, this analysis led to small discrepancies from what was predicted from the complete data set. In each case, changes were found in the lowest-frequency haplotypes (table 2). The efficacy of EMHAPFRE is known to decay as data sets decrease in size (Excoffier and Slatkin 1995). Thus, a second approach to ascription of haplotypes and their frequencies to each ethnic group was taken. To this end, a simple script was written in Microsoft Excel Visual Basic. This script, named "Assign," takes a list of haplotypes and a data set of unresolved genotypes and then assigns to each individual sample one or more pairs of haplotypes that can resolve its genotype data; Assign lists every pair of haplotypes that can resolve an individual's genotype data. We input each ethnic group's data set individually with the 22 haplotypes. In this way we were able to determine which of the haplotypes suggested by EMHAPFRE were necessary for resolution of our genotype data, thus further refining the results. The genotype of every sample in this study could be accounted for by at least one pair of the 22 haplotypes predicted by EMHAPFRE from the complete data set. Six of the 22 haplotypes exist in all ethnic populations, and 11 of them are unique to a single population and hereafter are referred to as "private" haplotypes (table 2); each of these 11 haplotypes has a frequency of <1%.

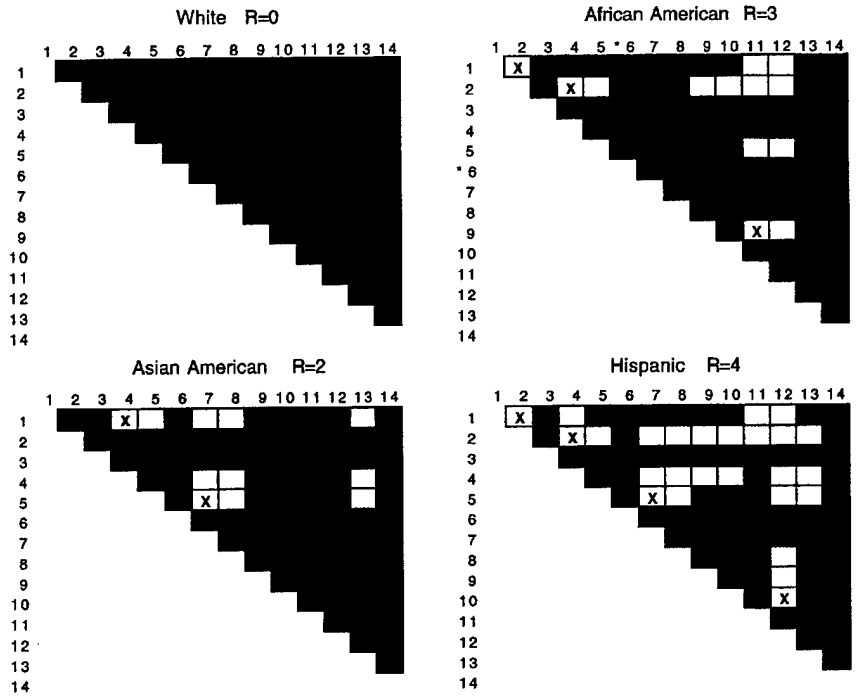


Figure 3 Four-gamete test for recombination in *ATM*. White boxes denote site pairs having four gametic types, which implies that recombination has occurred between these two sites. Also shown is the Hudson and Kaplan recombination statistic R, which is an estimate of the number and sites of recombination events needed to explain the results of the four-gamete matrix. A white box containing an " × " denotes a potential site of recombination. The asterisk (*) denotes an SNP that is not polymorphic in the sample population.

We analyzed primate DNA in order to approximate an ancestral *ATM* haplotype. Three haplotypes were found in 12 chromosomes (table 2). Two common chimpanzees, one bonobo, and three gorillas were genotyped by ASO hybridization and fluorescent sequencing; in cases in which ASO hybridization gave ambiguous results, fluorescent sequencing was used to confirm the genotype. None of the ape haplotypes was found among the 22 human haplotypes. One ape haplotype differs from a human haplotype by a single-base variant. This human haplotype is one of the least common (frequency .007) and occurs only in our African American study group. Only one of the human SNPs showed variation in the apes; the remainder were monomorphic. One common chimpanzee was heterozygous for IVS62+424g→a. The gorillas shared all but one allele with the chimpanzees. At IVS8−356t→c, gorillas are homozygous for a third allele (A), which is not found in either humans or chimpanzees.

Intragenic Recombination and LD

The small number of haplotypes seen in our study population suggests the possibility that recombination is reduced at the *ATM* locus. This is further evidenced by the results of the four-gamete test (fig. 3) (Hudson and Kaplan 1985). For a given haplotype AB, mutation may result in either Ab or aB. Haplotype ab arises only in the case of either recombination or repeat mutation. For the purpose of this analysis, we will consider repeat mutation to be rare and will use the four-gamete test as a measure of recombination. The four-gamete test was executed on unphased genotype data, in a pairwise fashion across SNP loci. This was done for each ethnic group separately. Interestingly, the four-gamete test found no site pairs with four gametes in the samples from white European Americans, implying a complete lack of recombination in that population. Low recombination was indicated for the other groups, as shown in figure 3.

Another test for recombination is that of Hudson and Kaplan (1985). Based on the resulting matrix of the four-gamete test, the Hudson and Kaplan parameter R is an estimate of the minimum number of recombination events in the history of the sample. For the white European American population, this estimate is 0 (fig. 3). For the other ethnic groups, R ranges from 4, in His-

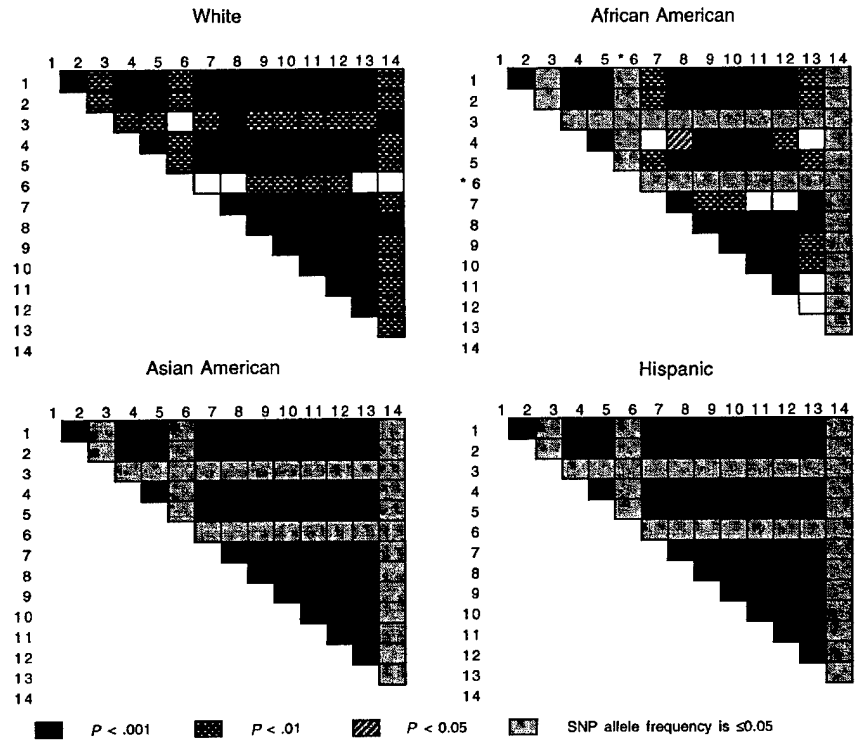


Figure 4 Fisher's exact test for LD in *ATM*. White boxes denote site pairs that do not have a significant value by Fisher's exact test, indicating linkage equilibrium. Gray columns and rows denote SNPs that have a minor-allele frequency ≤.05. The asterisk (*) denotes an SNP that is not polymorphic in the sample population.

panics, to 2, in Asian Americans. The predicted sites of recombination are similar among ethnic groups. African Americans and Hispanics share two possible recombination sites in the 5′ end of the gene, and a third, in the 3′ end, could also be in the same location. The Asian American population shares one of the 5′ end sites and has, in the middle of the gene, another potential site of recombination, which is also present in Hispanics.

Further support for the hypothesis that there is minimal recombination at the *ATM* locus is provided by the results of Fisher's exact test (Weir 1996). We computed all possible pairwise comparisons between sites, to determine the degree of nonrandom association between sites. The majority of site pairs across all data sets show significance ($P < .001$), indicating that there is extensive disequilibrium at this locus (fig. 4). It has been demonstrated that alleles with frequencies ≤.05 do not have the power for detection of disequilibrium (Lewontin 1995; Goddard et al. 2000). In this analysis, we included only SNPs having an allele frequency >.05. The Hispanic and Asian American populations were in complete disequilibrium. In the white European American population, the pattern of equilibrium followed the SNP with the lowest-frequency (.06) allele.

Disequilibrium was next measured by use of the statistic D', in a pairwise fashion across the 14 SNP loci (fig. 5). $D' = D/|D|\text{max}$, where $D = \text{p11} - \text{p1p2}$ and $|D|\text{max} = \text{max}(\text{p1p2},\text{q1q2})$ if $D < 0$ and $|D|\text{max} = \text{min}(\text{q1p2},\text{p1q2})$ if $D > 0$. D' ranges from 1 to −1, and this range is not influenced by allele frequency. A score of either 1 or −1 is considered to represent perfect disequilibrium. Interestingly, the results of this test are virtually superimposable on the results of the four-gamete test. The majority of site pairs are in perfect disequilibrium. The white European American population is in perfect disequilibrium across all sites. For the other groups, the sites with $|D'| < 1$ are exactly the same sites that have four gametes. We conclude that the *ATM* locus exhibits reduced recombination and extensive disequilibrium in all four ethnic groups, with the white European American population being the most extreme case.

Association Study

Ultimately, we aim to use these *ATM* haplotypes for association studies in populations with cancer. To evaluate the potential that these haplotypes have for identification of a particular mutation or polymorphism, we

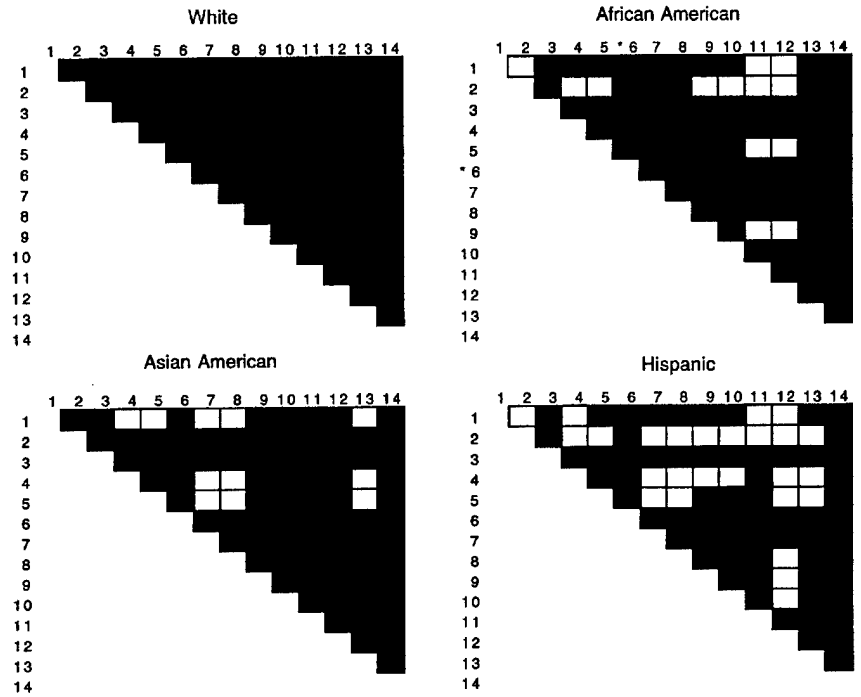


Figure 5 *D'*, measured as $D' = D/|D|\max$ in a pairwise fashion across 14 SNP loci. A score of either 1 or −1 is considered perfect disequilibrium. Black boxes denote site pairs with perfect disequilibrium; white boxes denote site pairs with $|D'| < 1$. The asterisk (*) pair denotes an SNP that is not polymorphic in the sample population.

performed a model association study. We tested the ability of these haplotypes to detect, by association, three different cSNPs in the *ATM* gene. These cSNPs were found by sequencing the reverse transcriptase–PCR products from *ATM* mRNA isolated from peripheral blood lymphocytes from cancer patients. cSNP1 is located in exon 4 and results in Ser49Cys. Positioned in exon 38, cSNP2 results in Asp1853Asn; and cSNP3, which results in Pro1054Arg, is located in exon 23. A population of 941 individuals was screened for these three cSNPs, by ASO hybridization. The resulting frequencies of the cSNPs in this population are shown in table 3.

In the model association study, samples from white European Americans in the 941-individual collection that were found to possess one of the three cSNPs were considered to be the "case" population. The "control" population consisted of samples from white European Americans from the same collection that were randomly chosen and negative for the cSNPs; because of the low frequency of these cSNPs in other ethnic groups, only samples from white European Americans were used in this association study. All case and control samples were genotyped for the 14 *ATM* neutral sequence variants, via ASO hybridization. To assign haplotypes to individual samples, we used Assign and the initial 22 *ATM* haplotypes.

Each cSNP showed a significant association with a different, specific *ATM* haplotype (table 4). cSNP1 showed an association with haplotype 2, cSNP2 with haplotype 15, and cSNP3 with haplotype 17. Haplotype 2 was present at a frequency of .29 in the control population (no. of chromosomes [c] = 152) and at a frequency of .64 in the cSNP1 population (c = 14); haplotype 15 was present at a frequency of .07 in the control population (c = 112) and at a frequency of .57 in the cSNP2 population (c = 56); and haplotype 17 was pre-

Table 3

Frequencies for Three *ATM* cSNPs in the Control Population

	Frequency in 941 Individuals			
	White European Americans	African Americans	Asian Americans	Hispanic Americans
cSNP1	.005	.000	.001	.001
cSNP2	.066	.001	.002	.017
cSNP3	.015	.001	.000	.005

Table 4

Association of *ATM* Haplotypes and *ATM* cSNPs in Individual "Case" and Control Populations

		Frequency in[a]		*P* for	
	Haplotype	Control Population	"Case" Population	Genotype Association[b]	Haplotype Association[c]
cSNP1	2	.29 (c = 152)	.64 (c = 14)	.0478	.0166
cSNP2	15	.07 (c = 112)	.57 (c = 56)	.0000	.0000
cSNP3	17	.08 (c = 146)	.52 (c = 54)	.0000	.0000

Note.—Samples carrying one of three *ATM* cSNPs were genotyped, by ASO hybridization, for the 14 *ATM* noncoding SNPs.

[a] Each cSNP was found to occur on separate *ATM* haplotypes. c = total number of white European American chromosomes genotyped.

[b] By 2 × 2 contingency table.

[c] By 3 × 3 contingency table.

sent at a frequency of .08 in the control population (c = 146) and at a frequency of .52 in the cSNP3 population (c = 54). These are 2-fold, 8-fold, and 6.5-fold increases in the frequencies of haplotypes 2, 15, and 17, respectively; and the *P* values for these associations are .0166, .0000, and .0000, respectively. Genotype correlations were also present, with *P* values of .0478, .0000, and .0000, respectively (table 4).

One of the great challenges in studying the genetics of a complex disease such as cancer is its multifactorial etiology. As presented in table 4, the data from our simulated association study model a scenario in which all "cases" are caused by a single mutation. To more accurately simulate an association study with a complex disease, we reanalyzed our data. We considered the three groups of samples carrying the variant cSNPs as one "case" population. In this analysis, two of the three haplotypes that originally had shown an association demonstrated a significant increase in frequency (table 5). No increase in frequency was apparent for haplotype 2, which had previously shown a twofold increase in the cSNP1 population; haplotype 15 showed a fourfold increase (P = .0002); and haplotype 17 showed a threefold increase (P = .0002). Thus, we successfully demonstrated the ability of these *ATM* haplotypes to discern members of our case population who carry a particular SNP. The results of these studies indicate a significant potential for the use of haplotypes extending over a large genomic region, to detect disease associations through a case-control–study design in a general population.

Discussion

In this study, we have presented a strategy for uncovering the genetic contribution to complex disease. Specifically, we have demonstrated the utility of a complex SNP-based–haplotype approach to association studies and have detected significant LD at the *ATM* locus, extending ~142 kb. The results of this study provide proof of principle for the use of SNP-haplotype data in the detection of genetic factors contributing to complex disease.

We sequenced 13.5 kb of the *ATM* gene in five unrelated individuals and detected 17 SNPs in noncoding regions. We then utilized these neutral sequence variants spanning 142 kb of the *ATM* gene to construct haplotypes for this genomic locus. The expectation-maximization algorithm EMHAPFRE (Excoffier and Slatkin 1995) was used to predict haplotypes from genotype data on 295 individuals from four ethnic groups. Twenty-two haplotypes and their frequencies were predicted by EMHAPFRE, for the total population. Three of these 22 haplotypes have a frequency of ⩾10%. This concurs with the findings of Li et al. (1999), who also used neutral sequence variants to detect three major haplotypes at the *ATM* locus. Six of the 22 haplotypes exist in all four ethnic populations in our study and are also the most commonly occurring haplotypes. There are 11 private haplotypes, each of which has a frequency of <1%.

We verified the reliability of the haplotype-prediction algorithm by using several tests. First, we genotyped individuals from nine three-generation CEPH families (n = 87). This allowed us to determine haplotypes by inspection of allele segregation. The CEPH genotype data were also analyzed by EMHAPFRE, and the resulting haplotypes agreed completely with those in-

Table 5

Association of *ATM* Haplotypes and *ATM* cSNPs in Combined Case Population

	Frequency in Combined Case Population	*P* for	
Haplotype	(c = 124)	Genotype Association	Haplotype Association
2	.19	.9644	.7606
15	.27	.0000	.0002
17	.24	.0000	.0002

Note.—See footnotes to table 4.

ferred on the basis of transmission data. Next, we used Assign, a script written in Microsoft Excel Visual Basic, to assign pairs of haplotypes to individual genotypes. Given the 22 haplotypes predicted by EMHAPFRE, Assign successfully resolved the genotype data for all 295 individuals in this study. The results of EMHAPFRE were tested against another haplotype-prediction program, one that does not use the expectation-maximization algorithm and that does not assume that Hardy-Weinberg is in effect. This program, termed "Data Mining," uses the resulting matrix of the four-gamete test to inform the process of haplotype prediction so that recombination may influence outcome (N. Wang, R. Chakraborty, M. Kimmel, and L. Jin, personal communication). There were minor differences in the results of this comparison. For the population of white European Americans, the outcome of each program was identical. This is not surprising, since the four-gamete test reveals no evidence for recombination in this population. The results of these trials confirm that EMHAPFRE was successful in estimating the correct haplotypes necessary to sufficiently resolve our data set. We feel confident that the size and diversity of our data set has allowed us to describe in relative depth the haplotype architecture of *ATM*. Consequently, we have chosen to use, as the foundation for further studies, the 22 haplotypes predicted from the complete data set.

With a minimal amount of sequencing (13.5 kb in five individuals), we were able to detect highly informative neutral sequence variants spanning a large genomic region. In sequencing 10 chromosomes form white European Americans, we found SNPs that have a common occurrence in four different ethnic groups. In all ethnic groups, the majority (11 of 14) of SNPs identified in this study have a minor-allele frequency of $\geq$25%. SNPs with frequencies in the range of .2–.5 have the highest information content for association and LD studies (Kruglyak 1997). Although most SNPs had a high minor-allele frequency in all ethnic groups, allele frequencies varied across ethnic groups. This is in accordance with several other studies that have found population differences in SNP-allele frequencies (Lai et al. 1998; Nickerson et al. 1998; Cargill et al. 1999; Halushka et al. 1999; Goddard et al. 2000). Variations in allele frequencies are most pronounced in the African American population. Four SNPs (IVS21−77t→c, IVS55+186c→t, IVS62+424g→a, and IVS62−973a→c) have a minor-allele frequency that is reduced by 40%–75% in African Americans, compared with that in other ethnic groups. A fifth SNP, IVS46−257a→c, was not found in the African American samples. These differences illustrate that there is population structure in SNP-allele frequencies that is an important factor to consider when SNP-based association and LD studies are designed.

Comparison of genotype data from six great apes was instructive for approximating ancestral haplotypes and SNP alleles. Genotyping revealed three haplotypes in this population, none of which is identical to the human *ATM* haplotypes. Of the 14 SNPs, 2 showed variation in the ape population. One common chimpanzee was heterozygous for IVS62+424g→a, and all three gorillas were homozygous for a third allele (A) at IVS8−356t→c. The extent of homozygosity in this sample indicates that most of the SNPs found varying in the human population have arisen since the divergence of the human lineage from the last common ancestor shared with the chimpanzee. This agrees with the assertion by Hacia et al. (1999)—that is, that most current neutral human polymorphisms are not shared with the chimpanzee (Hacia et al. 1999). It may also imply that these SNPs are not hypermutable sites, since more variation might be expected in the 12 primate chromosomes analyzed. Although these SNPs are common in man, they are not due to hypermutability; rather, they are old enough to be found throughout diverse ethnic groups.

The results of this study show a remarkable lack of recombination at the *ATM* locus. This effect is most profound in the white European American population, in which no evidence for recombination is detected by the four-gamete test and in which D' shows perfect disequilibrium across all SNPs. Low recombination is implicated for the African American, Asian American, and Hispanic groups as well. The possibility of low recombination was suspected on the basis of the seemingly small number of haplotypes found at this locus. Twenty-two haplotypes with 14 loci is not considerably greater than the $n + 1$ (i.e., 15) that would be expected if there is no recombination. Another study, performed in parallel with this one, used the same approach as that described here and serves as a direct comparison: D. Trikka, Z. Fang, A. Renwick, S. Jones, R. Chakraborty, M. Kimmel, and D. L. Nelson (unpublished data) used neutral sequence variants dispersed across the BLM, WRN, and RECQL loci, to derive haplotypes for these regions; their study used the same sample population, with fewer SNPs (8, 13, and 11 respectively) for haplotype construction, and found considerably larger numbers of haplotypes (50, 56, and 47, respectively) at each locus. The key difference between these loci and *ATM* is the amount of recombination and LD reported. Trikka et al. found more evidence for recombination and linkage equilibrium when the four-gamete test and Fisher's exact test were used. For *ATM,* the four-gamete test revealed few site pairs with four gametes. The Hudson-Kaplan recombination statistic R ranged from 0, in white European Americans, to 4, in Hispanics. Analysis by both Fisher's exact test and D' indicated extensive LD for *ATM,* in all ethnic groups studied. Figure 5

shows extensive disequilibrium, with >72% of site pairs having perfect disequilibrium in all ethnic groups.

Using a model association study, we have successfully demonstrated the ability of *ATM* haplotypes to identify chromosomes carrying specific coding polymorphisms. The three cSNPs that we used as candidates for detection had varying frequencies in our control population of white European Americans (cSNP1, .005; cSNP2, .066; and cSNP3, .015). When each of the three cSNP populations was analyzed individually, each cSNP showed a significant association with a different *ATM* haplotype. cSNP1 showed an association with haplotype 2, cSNP2 with haplotype 15, and cSNP3 with haplotype 17 (P = .0166, .0000, and .0000, respectively); the increase in haplotype frequency in cases versus controls was 2-fold, 8-fold, and 6.5-fold, respectively. To model the potential for multiallelic etiology of a complex disease, we combined the three populations of samples carrying the cSNPs into one "case" population. In this analysis, two haplotypes demonstrated a readily detectable increase in frequency: haplotype 15 showed a fourfold increase, and haplotype 17 showed a threefold increase in frequency; no frequency increase was apparent for haplotype 2, which had previously shown a twofold increase in the cSNP1 population.

The association that becomes undetectable (i.e., haplotype 2 with cSNP1) involves the haplotype occurring most commonly (frequency .29) in the general population. Haplotype 15 shows the greatest increase in frequency and is the least common of the three haplotypes, with a control frequency of .05. This leads us to an important point for future association studies. Haplotypes with lower frequencies in control populations may be more effective for detection of associations. However, it is important to note that haplotype 17, which is the third most frequent haplotype (frequency .10), nevertheless showed a 2.6-fold increase in frequency in the combined cSNP population. An additional factor contributing to detection in this study is frequency of the mutation. In the case of cSNP1 and haplotype 2, in which the association becomes undetectable, the most frequent haplotype was associated with the least common SNP (cSNP1, .006). The difference in frequency between cSNP1 and the other cSNPs is a factor of 10. Both the haplotype frequency and the cSNP frequency contribute to detection. This underscores the idea that several factors, including frequency of haplotype, frequency of mutation, and age of mutation, contribute to limits of detectability.

This model association study demonstrates proof of principle for the use of complex SNP haplotypes covering candidate genes, in the detection of genetic factors contributing to complex disease. We have successfully demonstrated the ability of these *ATM* haplotypes to discern members of our "case" population that carry a particular coding SNP. The results of these studies indicate that haplotypes extending over a large genomic region have a significant potential for detection of disease associations.

There is much interest in the use of SNPs in genomewide association studies and other LD-based strategies. Our approach and analyses bear on those strategies, in several regards. First, LD estimates from simulation studies have been as low as 3 kb of meaningful LD (Kruglyak 1999). This calculation suggests that a very-high-density map with as many as 0.5–3 million SNPs would be necessary for effective association studies (Kruglyak 1999). Our results and those of other studies (Collins et al. 1999; Eaves et al. 2000; Moffatt et al. 2000; Taillon-Miller et al. 2000) indicate, to the contrary, that significant LD can be found extending as far as several hundred kilobases. This should reduce the number of SNPs necessary for genomewide linkage studies. Comparison of LD at *ATM* versus the results of the *LPL* study (Clark et al. 1998) in which LD patterns were complex over just 9.7 kb supports the idea that LD varies widely throughout the genome, indicating that some regions will require SNPs that are more densely spaced. Second, higher-frequency (.2–.5) SNPs are more robust, whereas rare SNPs may be less useful and, in some analyses, may confound results. More than half of the SNPs used to construct haplotypes in the *LPL* study had a relative allele frequency of <.2. This resulted in 67 of 71 individuals having a unique haplotype. By using fewer markers (14) with higher frequency (.20), we were able to effectively elucidate the haplotype architecture and the LD and recombination profiles for the *ATM* genomic locus (142 kb). These haplotypes were used successfully in association studies, to detect coding polymorphisms in the *ATM* gene. We conclude that reasonably spaced, highly informative SNPs have the ability to define a larger number of ancestral chromosomes and have increased power for population-based association studies.

Acknowledgments

The authors thank Ranajit Chakraborty of the University of Texas Health Science Center (Houston) and Marek Kimmel of Rice University (Houston), for discussion and review of the manuscript; Jason Deats, for his valued contribution to the design and code for Assign; and Melissa Bondy, Alice Sigurdson, and Scott Manatt, of M.D. Anderson Cancer Center, for collection of normal DNAs. P.E.B was supported by a fellowship from the W.M. Keck Center for Computational Biology (Houston) (funded by U.S. National Library of Medicine Training Grant 1T15LM07093). T.A.B. was supported by Department of Defense Breast Cancer Research Program Career Development Award CBC980154. This work was supported by the Kleburg Fund for New and Innovative Research and by U.S. National Cancer Institute grant CA75432.

Appendix A

Primers Used for PCR and Sequencing

f1.atm, ATGGTCATCTCGTTACAGGCAATGC
r1.atm, CCCCAAGTGACTGAAGGCATCTAGG
f2.atm, TGGTGGAACCATTTCCGTTTAACG
r2.atm, GCGCCCTTCTAATAACCCGCC
f3.atm, GCCCAGAACCTCCGAATGACG
r3.atm, CGACTTAGCGTTTGCGGCTCG
f4.atm, TGGCTGGCAACATTACCAACTGC
r4.atm, TGCATCTTTTTCTGCCTGGAGGC
f5.atm, TGTGTGCTAGGGAGGAATCTGGTGG
r5.atm, GGCTGTCTCTAGGCTTGTTGAGGGC
f6.atm, CCATCATCCGAAAGGAGCCAAAAC
r6.atm, GCAGCAATTTCCCTGTTTCTGCC
f7.atm, AAATTGGCAGGATGATGAGGATGC
r7.atm, GCTGTCAAGCTGCATCAGCGTTAG
f8.atm, CCAAAGCGTGCCAGAATGGTATG
r8.atm, CCAAAGCGTGCCAGAATGGTATG
f9.atm, GGTATGCGTAGCGGGGCTAGTGAG
r9.atm, CGCAGGAAAAAGCCAGATGCAATC
f10.atm, GCCCTAGCCCCAGTGTATGTGGAG
r10.atm, GGCAGCCAGTTTCCGAGAACTACC
f11.atm, TTTTTGGCAAGGTGAGTATGTTGGC
r11.atm, TGCGAACTTGGTGATGATTGTCAGC
f12.atm, AGATTGTTCCAGGACACGAAGGGAG
r12.atm, TTTCTTCCCATTGTCACCTGTTCCC
f13.atm, TGCGAAAAACAGGCTTTGTTTGC
r13.atm, GGTGATGGAAAAGAGACGGGGC
f14.atm, GCAAGTCCCCTCACCAGCAACAC
r14.atm, GATGCCTTCCCATCATCCTGATACC
f15.atm, TCTGGGAAGAAGTTACGCAGGGAAC
r15.atm, CTGACTGGCACTAGAATTTGCTGGC
f16.atm, GGCGGAATGAATGTGAGTTATGCG
r16.atm, CCAGGTGATTTCTCCATCCCGTG
f17.atm, CTGCCTAAAGCAGCAGTTTTTGCC
r17.atm, TGTTGCTATCCCGAAGCTGAAACC
f18.atm, GGTGTGTAAGCAAGAATGCCTGGG
r18.atm, GCCACAGATTTTGAGACCACTGCAC
f19.atm, TAGTTTGTATGGCTGTGGTGGAGGG
r19.atm, CATCCCTCTGCTTCAGGAGTATCCC
f20.atm, CCAGTAGGGGGTCCCTCATTTCC
r20.atm, TGAGAAGCTGGGAGTGTTTCTGCC
f21.atm, CCCCGTACATGAAGGGCAGTTG
r21.atm, TGGGTGGCTGGGCTAATGAAGAG
f22.atm, GGTTCAGCGAGAGCTGGAGTTGG
r22.atm, GCAGCAGGGGGAAAACCCAC
f23.atm, CCACAGATTAGCAACAAGTTGGGGC
r23.atm, TGGCATAAGCACACGGAAACTCTCC
f24.atm, AGGTTCCGATGGCAAGGAGAGG
r24.atm, CTGTGTCTTTCCACCACTCCCCAG
f25.atm, CAGTCATGGTTCTGGGGAGAGAAGC
r25.atm, GCCTTTCTGATTTCCCTTCCTGCC
f26.atm, CTTGATGGTGGGAGGGACTTAGGG
r26.atm, TGCCTAGATGTTTGAGAGCCTGCC
f27.atm, CAGGGCACACAGGGTACAGTGTAGG
r27.atm, TCAGTTCAGACCATCTCATGCCTCC
f28.atm, CAGGGGGATGATAGTGATGATGTGG
r28.atm, TTCAAAACATACATGCCCTGCCTTC
f29.atm, CAAAGACTGAGAGCTGAGCCCAGTG
r29.atm, GCACAATCTCCTCCTTTCTGCTGC
f30.atm, TGGTTTAGAAATGCCTTCAGCCCC
r30.atm, TGCACTCTACCTGCCATGCTTCC
f31.atm, GCCATGTCAGTGCCCAACTTGAAG
r31.atm, TTGGTGCTGCGTTTGGAATCTTG
f32.atm, GATTCCAAACGCAGCACCAAACC
r32.atm, GGTAGTTGATGGGGGAGGGGAAC
f33.atm, GTTCCCCTCCCCCATCAACTACC
r33.atm, GAGCACAGTGCCTTCTTCCACTCC
f34.atm, CCCTGACAATCTGGGGCACAAAC
r34.atm, CCGTGGCTTTTGCTGGCATTC
f35.atm, GTCCTGTGGCATTGTGCATAACTCC
r35.atm, GCAGACATTAGGCATAAGCCCCTTC
f36.atm, GATGACTGCCCTTGTTCCCCAAG
r36.atm, TGGTTAAGTTGCTTTTCCCCCCAG

Appendix B

Primer Sequences and Concentrations Used for Multiplex PCR

Group 8:
3F *ATM*, 5′-GCCCAGAACCTCCGAATGACG-3′; and 3R-2 *ATM*, 5′-GCCGTGAAGCGAAAGAGGCG-3′ (0.25 μM)
11F *ATM*,5′-TTTTTGGCAAGGTGAGTATGTTGGC-3′; and 11R *ATM*, 5′-TGCGAACTTGGTGATGATTGTCAGC-3′ (0.25 μM)
14F *ATM*, 5′-GCAAGTCCCCTCACCAGCAACAC-3′; and 14R *ATM*, 5′-GATGCCTTCCCATCATCCTGATACC-3′ (0.25 μM)
23F-2 *ATM*, 5′-GGTGGAATCTGGTCTAGTTACCC-3′; and 23R *ATM*, 5′-TGGCATAAGCACACGGAAACTCTCC-3′ (0.25 μM)
27F *ATM*, 5′-CAGGGCACACAGGGTACAGTGTAGG-3′; and 27R *ATM*, 5′-TCAGTTCAGACCATCTCATGCCTCC-3′ (0.188 μM)
29R *ATM*, 5′-CAAAGACTGAGAGCTGAGCCCAGTG-3′; and 29R *ATM*, 5′-GCACAATCTCCTCCTTTCTGCTGC-3′ (0.125 μM)
30F *ATM*, 5′-TGGTTTAGAAATGCCTTCAGCCCC-3′; and 30R-2 *ATM*, 5′-CAGCCAGTCCAACATAAATCAG-3′ (0.25 μM)
31F *ATM*, 5′-GCCATGTCAGTGCCCAACTTGAAG-3′; and 31R *ATM*, 5′-TTGGTGCTGCGTTTGGAATCTTG-3′ (0.25 μM)
Group 7:
7R *ATM*, 5′-GCTGTCAAGCTGCATCAGCGTTAG-3′; and 7F-2 *ATM*, 5′-GTTGGATTACCATGTTCACCAG-3′ (.188 μM)
10F *ATM*, 5′-GCCCTAGCCCCAGTGTATGTGGAG-3′and 10R-2 *ATM*, 5′-GCAGAGATAATCATGGGCAGG-3′ (0.25 μM)
15F *ATM*, 5′-TCTGGGAAGAAGTTACGCAGGGAAC-3′; and 15R-2 *ATM*, 5′-TGGGGAGACTATGGTAAAAGAGG-3′ (0.31 μM)
16F *ATM*, 5′-GGCGGAATGAATGTGAGTTATGCG-3′; and 16R *ATM*, 5′-CCAGGTGATTTCTCCATCCCGTG-3′ (0.25 μM)
20F *ATM*, 5′-CCAGTAGGGGGTCCCTCATTTCC-3′; and 20R *ATM*, 5′-TGAGAAGCTGGGAGTGTTTCTGCC-3′ (0.25 μM)

25F *ATM*, 5′-CAGTCATGGTTCTGGGGAGAGAAGC-3′; and 25R-2 *ATM*, 5′-CTATCAATATCTAGCTCTGGGGC-3′ (0.15 μM)
28F *ATM*, 5′-CAGGGGGATGATAGTGATGATGTGG-3′; and 28R *ATM*, 5′-TTCAAAACATACATGCCCTGCCTTC-3′ (0.5 μM)

Appendix C

Probes Used for ASO Hybridization

*ATM*Aso 3T, 5′-TAACCCTCCTTCCCGC-3′
*ATM*Aso 3a, 5′-TAACCCTCCATCCCGC-3′
*ATM*Aso 7T, 5′-AAGGAACTTGTAATATTTTTC-3′
*ATM*Aso 7c, 5′-AGGAACTCGTAATATTTTTC-3′
*ATM*Aso 10T, 5′-TGGGAAACATGACCAGGG-3′
*ATM*Aso 10c, 5′-GGGAAACACGACCAGGG-3′
*ATM*Aso 11T, 5′-GTAACTTATAATAACCTTTC-3′
*ATM*Aso 11c, 5′-GAAGTAACTTACAATAACC-3′
*ATM*Aso 14C, 5′-TCTGTACAAGAAAAATTTG -3′
*ATM*Aso 14g, 5′-TCTGTAGAAGAAAAATTTG-3′
*ATM*Aso 15C, 5′-TTTCCTCTCAGTCTACAGG-3′
*ATM*Aso 15t, 5′-TTTTTCCTCTTAGTCTACAGG-3′
*ATM*Aso 16C, 5′-TAGAGATGATGTCGGCTTC-3′
*ATM*Aso 16t, 5′-CTAGAGATGATGTTGGCTTC-3′
*ATM*Aso 20A, 5′-GTAATGTCAGAGTATTAAA-3′
*ATM*Aso 20c, 5′-TAATGTCAGCGTATTAAA-3′
*ATM*Aso 23T, 5′-CAAAAGCTTCTCTTGCTTT-3′
*ATM*Aso 23c, 5′-AAAAGCTTCTCCTGCTTTC-3′
*ATM*Aso 25C, 5′-TTTTTTGTGGCATTCACAC-3′
*ATM*Aso 25t, 5′-TTTTTTGTGGTATTCACAC-3′
*ATM*Aso 27C, 5′-CTGCTCATGCCTCCTCTC-3′
*ATM*Aso 27t, 5′-CTGCTCATGTCTCCTCTCC-3′
*ATM*Aso 28C, 5′-TTCTATTAAACAGTATTA-3′
*ATM*Aso 28a, 5′-TTCTATTAAAAAGTATTA-3′
*ATM*Aso 29.1T, 5′-GATAAAGATATGTTGACAA-3′
*ATM*Aso 29.1c, 5′-GATAAAGATACGTTGACAA-3′
*ATM*Aso 29.2C, 5′-ACTTCCTGACGAGATACAC-3′
*ATM*Aso 29.2t, 5′-ACTTCCTGATGAGATACAC-3′
*ATM*Aso 30c, 5′-CCTAAGCCACGTTCCTCTA-3′
*ATM*Aso 30t, 5′-CCTAAGCCATGTTCCTCTA-3′
*ATM*Aso 31.1C, 5′-AAATAGAGCGATTTTGGTT-3′
*ATM*Aso 31.1t, 5′-AAATAGAGAGATTTTGGTTC-3′
*ATM*Aso 31.2C, 5′-AGAAATTCCTCATGAACTC-3′
*ATM*Aso 31.2a, 5′-AGAAATTCATCATGAACTC-3′

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank Overview, http://www.ncbi.nlm.nih.gov/Genbank Overview.html (for genomic sequence [accession number U82828])

References

Boerwinkle E, Ellsworth DL, Hallman DM, Biddinger A (1996) Genetic analysis of atherosclerosis: a research paradigm for the common chronic diseases. Hum Mol Genet 5:1405–1410

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238 (erratum: Nat Genet 23:373 [1999])

Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF (1998) Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am J Hum Genet 63:595–612

Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 96:15173–15177

Collins FS, Guyer MS, Charkravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278:1580–1581

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R (1990) Centre d'Étude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. Genomics 6:575–577

DeMarchi JM, Richards CS, Fenwick RG, Pace R, Beaudet AL (1994) A robotics-assisted procedure for large scale cystic fibrosis mutation analysis. Hum Mutat 4:281–290

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA (2000) The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat Genet 25:320–323

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927

Gatti RA, Boder E, Vinters HV, Sparkes RS, Norman A, Lange K (1991) Ataxia-telangiectasia: an interdisciplinary approach to pathogenesis. Medicine (Baltimore) 70:99–117

Gatti RA, Tward A, Concannon P (1999) Cancer risk in ATM heterozygotes: a model of phenotypic and mechanistic differences between missense and truncating mutations. Mol Genet Metab 68:419–423

Goddard KA, Hopkins PJ, Hall JM, Witte JS (2000) Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. Am J Hum Genet 66:216–234

Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, Collins FS (1999) Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet 22:164–167

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of

single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22:239–247

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Josefsson AM, Magnusson PK, Ylitalo N, Quarforth-Tubbin P, Ponten J, Adami HO, Gyllensten UB (1998) p53 polymorphism and risk of cervical cancer. Nature 396:531–532

Kraft HG, Windegger M, Menzel HJ, Utermann G (1998) Significant impact of the +93 C/T polymorphism in the apolipoprotein(a) gene on Lp(a) concentrations in Africans but not in Caucasians: confounding effect of linkage disequilibrium. Hum Mol Genet 7:257–264

Kruglyak L (1997) The use of a genetic map of biallelic markers in linkage studies. Nat Genet 17:21–24

——— (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22: 139–144

Lai E, Riley J, Purvis I, Roses A (1998) A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. Genomics 54:31–38

Lander ES (1996) The new genomics: global views of biology. Science 274:536–539

Lewontin RC (1995) The detection of linkage disequilibrium in molecular sequence data. Genetics 140:377–388

Li A, Huang Y, Swift M (1999) Neutral sequence variants and haplotypes at the 150 Kb ataxia-telangiectasia locus. Am J Med Genet 86:140–144

MacDonald ME, Novelletto A, Lin C, Tagle D, Barnes G, Bates G, Taylor S, Allitto B, Altherr M, Myers R, Lehrach H, Collins F, Wasmuth J, Frontali M, Gusella J (1992) The Huntington's disease candidate region exhibits many different haplotypes. Nat Genet 1:99–103

Moffatt MF, Traherne JA, Abecasis GR, Cookson WO (2000) Single nucleotide polymorphism and linkage disequilibrium within the TCR alpha/delta locus. Hum Mol Genet 9:1011–1019

Morrell D, Chase CL, Swift M (1990) Cancers in 44 families with ataxia-telangiectasia. Cancer Genet Cytogenet 50:119–123

Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 19:233–240

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847–856

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Rotman G, Shiloh Y (1999) ATM: a mediator of multiple responses to genotoxic stress. Oncogene 18:6135–6144

Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle DA, et al (1995) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 268:1749–1753

Stankovic T, Kidd AM, Sutcliffe A, McGuire GM, Robinson P, Weber P, Bedenham T, Bradwell AR, Easton DF, Lennox GG, Haites N, Byrd PJ, Taylor AM (1998) ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer. Am J Hum Genet 62: 334–345

Storey A, Thomas M, Kalita A, Harwood C, Gardiol D, Mantovani F, Breuer J, Leigh IM, Matlashewski G, Banks L (1998) Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. Nature 393:229–234

Strittmatter WJ, Roses AD (1995) Apolipoprotein E and Alzheimer disease. Proc Natl Acad Sci USA 92:4725–4727

Swift M, Morrell D, Cromartie E, Chamberlin AR, Skolnick MH, Bishop DT (1986) The incidence and gene frequency of ataxia-telangiectasia in the United States. Am J Hum Genet 39:573–583

Swift M, Morrell D, Massey RB, Chase CL (1991) Incidence of cancer in 161 families affected by ataxia-telangiectasia. N Engl J Med 325:1831–1836

Swift M, Reitnauer PJ, Morrell D, Chase CL (1987) Breast and other cancers in families with ataxia-telangiectasia. N Engl J Med 316:1289–1294

Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice JP, Kwok PY (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat Genet 25:324–328

Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland, MA

Yu CE, Oshima J, Fu YH, Wijsman EM, Hisama F, Alisch R, Matthews S, Nakura J, Miki T, Ouais S, Martin GM, Mulligan J, Schellenberg GD (1996) Positional cloning of the Werner's syndrome gene. Science 272:258–262

RADIATION-INDUCED CHROMATID BREAKS AS A PREDICTOR OF BREAST CANCER RISK

Thomas A. Buchholz, M.D.[1], Xifeng Wu, M.D., Ph.D.[2]

Departments of [1]Radiation Oncology and [2]Epidemiology,

The University of Texas M. D. Anderson Cancer Center, Houston, TX

Correspondence Address: Xifeng Wu, M.D., Ph.D.

Department of Epidemiology, Box 189

The University of Texas M. D. Anderson Cancer Center

1515 Holcombe Blvd.

Houston, TX 77030

Phone (713) 792-3020, fax (713) 792-0807, email:xwu@notes.mdacc.tmc.edu

Oral presentation, First International Conference on Translational Research and Pre-Clinical Strategies in Radio-Oncology, Lugano, Switzerland, March, 2000

This work was supported by T32CA77050 awarded by the National Cancer Institute, U.S. Department of Health and Human Services and T.A.B. is supported by a USAMRMC Career Development Award from the Breast Cancer Research Program (BC980154).

running head: radiosensitivity and breast cancer development

Abstract

Background/Purpose: In *in vivo* models, radiation-induced genomic instability correlates with the risk of breast cancer development. In addition, homozygous mutations in tumor suppressor genes associated with breast cancer development adversely affects the processing and repair of radiation-induced DNA damage. We performed a case-control study to determine whether an assay measuring radiation-induced chromatid breaks correlated with the risk of having bilateral breast cancer.

Materials & Methods: Patients were prospectively studied on an institutional review board-approved protocol. We included only women with bilateral breast cancer as cases in order to obtain patients with a presumed genetic susceptibility for breast cancer. Controls were healthy women without a previous cancer history. A mutagen sensitivity assay using γ-radiation was performed on lymphocytes obtained from 26 cases and 18 controls. One milliliter of whole blood was cultured with 9 ml of blood medium for 67 h and then treated with 125 cGy using a Cs-137 irradiator. Following an additional 4 h in culture, cells were treated with Colcemid for 1 h to arrest cells in metaphase. The number of chromatid breaks per cell was counted using a minimum of 50 metaphase spreads for each sample.

Results: Cases had a statistically higher number of γ-radiation-induced chromatid breaks per cell than controls, with mean values of 0.61 +/- 0.24 versus 0.45 +/- 0.14, respectively ($p=0.034$, Wilcoxon rank sum test). Using the 75th percentile value in the control group

as a definition of radiation sensitivity, the radiation-sensitive individuals had a 2.83-fold increased odds ratio for breast cancer development compared with individuals who were not radiation sensitive (95% confidence intervals of 0.83, 9.67).

Conclusions: These preliminary data suggest that sensitivity to radiation-induced chromatid breaks in lymphocytes correlates with the risk of bilateral breast cancer. Although the differences between cases and controls were statistically significant, the small sample size necessitates that this finding be validated in a larger study. More data are also needed to determine whether this sensitivity is limited to breast cancer patients with a genetic susceptibility for the disease or also applies to the general breast cancer population.

Key Words: breast cancer, radiosensitivity, chromatid breaks, radiation

Introduction

A biological predictive assay of breast cancer development risk would have significant relevance to a large cohort of women. Breast cancer is the most common nondermatological cancer in women, with an estimated 182,800 new cases diagnosed annually in the United States (1). Furthermore, breast cancer remains the second leading cause of cancer deaths in women, with 40,800 predicted to die of the disease in the year 2000 (1). A biological assay quantifying an individual's risk of breast cancer development could help identify candidates for judicious clinical and radiographic screening, for trials evaluating chemoprevention strategies, and for consideration of prophylactic surgical interventions.

There are two types of biological predictors of breast cancer risk: genotype sequencing and phenotype screening. The discovery and cloning of BRCA1 and BRCA2 have permitted the development of a commercially available sequencing test to identify germline mutations in these two genes. In addition, the relationships of mutations in other candidate genes, such as ATM (ataxia telangiectasia, mutated), to breast cancer development are being investigated by a number of groups. While genotype sequencing has had a dramatic impact on understanding breast cancer risk in selected cases, only a small percentage of breast cancer patients develop the disease in the setting of a known predisposing germline mutation (2,3).

In this preliminary report, we investigate a phenotype-screening assay for breast cancer. Phenotype screening has a number of advantages and disadvantages compared with genotype sequencing. By evaluating a common downstream consequence of a

variety of tumor suppressor gene mutations, a phenotype assay can potentially capture a much broader percentage of the breast cancer population. Furthermore, this strategy is not dependent on new gene discovery and potentially can identify individuals who harbor relevant germline mutations in yet undiscovered genes. A phenotype-screening assay also affords the possibility of quantifying the importance of an individual's genotype. For example, a phenotype-screening assay may be able to quantitatively distinguish between different mutations in a tumor suppressor gene that entail different risks of breast cancer development.

The assay we investigated in this study was cellular radiosensitivity, as defined by the number of chromatid breaks per cell following *in vitro* treatment of lymphocytes with γ-radiation.

Materials and Methods

Approval for this prospective study was obtained through The University of Texas M. D. Anderson Institutional Review Board. Informed consent was obtained from all cases and controls in this study.

Cases were 26 women with a history of bilateral breast cancer. All women had a least one breast cancer treated in our institution. No samples were obtained from the cases during chemotherapy or radiation treatment because these treatments could potentially affect the number of chromatid breaks. Controls were 18 women with no personal cancer history who were recruited for a simultaneous study investigating lung cancer.

All participants donated 10 - 20 ml of blood for the mutagen sensitivity assay. The details of the mutagen sensitivity assay have been previously described (4), although in this study γ-radiation rather than bleomycin was used a mutagenic agent. All cultures were set up within 24 h of the blood draw. One milliliter of blood was added to 9 ml of RPMI-1640 medium supplemented with 20% fetal calf serum and phytohemagglutinin. The cultures were then incubated at 37°C for 72 h, after which the cultures were treated with 125 cGy of γ-radiation delivered from a Cs-137 irradiator. The cultures were then incubated for 4 h to allow time for DNA repair. Subsequently, the cultures were treated for 1 h with Colcemid (0.04 μg/ml) to arrest cells in metaphase. Cells were then harvested, fixed, washed, and stained with Giemsa as previously reported (4). For each case and control, the number of chromatid breaks per cell were counted. A minimum of 50 metaphase spreads per sample were examined.

The mean values, standard deviations, and standard errors were calculated. A Wilcoxon rank sum test for non-normal distribution was used to compare cases and controls. This test analyzed the data as categorical variables to minimize the impact that a single high mutagen sensitivity score could have on the mean value. Odds ratios for bilateral breast cancer were determined by comparing the incidence of bilateral breast cancer in mutagen-sensitive individuals and mutagen-resistant individuals. Consistent with previous reports, the value for being categorized as mutagen sensitive was the 75% value of the control population. This value was determined prior to the analysis of the data and represents an accepted quartile cutoff point.

Results

Characteristics of Cases

The median age of the cases at the time of first breast cancer diagnosis was 49 years with a range of 25 – 79. Sixty-five percent of the cases had a history of breast cancer in either a primary relative (38%) or a secondary relative (27%), 31% of the cases denied a breast cancer family history, and in 1 case the family history was unknown. The majority of cases (88%) were Caucasian. Only two of the cases were of Ashkenazi Jewish decent, and 1 of these was known to have a germline mutation in BRCA1. From a published nomogram for predicting the probability of having a BRCA1 mutation based on personal cancer history, family cancer history, age of diagnosis, and whether the individual is of Ashkenazi descent (5), the approximate average probability of having a BRCA1 mutation for our cases was 15%.

Mutagen Sensitivity Assay Results

The number of chromatid breaks per cell was significantly higher in our cases versus controls, with respective values of 0.61 +/- 0.24 (standard deviation) and 0.45 +/- 0.14 ($p=0.034$). Figure 1 shows the distribution of cases and controls according to the number of chromatid breaks per cell. As shown, the distribution of the cases is skewed to the radiosensitive end of the graph.

The data were also analyzed to determine the odds ratio for breast cancer development for mutagen-sensitive and mutagen-resistant individuals. Consistent with previous studies using the mutagen sensitivity assay, we dichotomized cases and controls

as being sensitive or resistant at the 75% level of the controls (0.56 chromatid breaks per cell). This analysis revealed that the mutagen-sensitive individuals had an odds ratio for breast cancer development of 2.83 (95% confidence interval of 0.83 – 9.67).

A comparison was also performed between the cases with a positive (n=17) or negative (n=8) family history. These results revealed that the cases with a positive family history had a higher number of chromatid breaks per cell than those with a negative family history, although the difference between the two groups was not statistically significant (0.67 +/- 0.14 versus 0.49 +/- 0.25, p=0.07). The distribution of these results is shown in Figure 2.

Discussion

In this paper, we present evidence that the phenotype of cellular radiosensitivity, as defined by a chromatid-break assay, correlates with the risk of having bilateral breast cancer. Specifically, we found that radiation induced a greater number of chromatid breaks in lymphocytes from patients with a history of bilateral breast cancer compared to female controls without a cancer history.

This study followed an earlier negative report from our institution investigating the value of the mutagen sensitivity assay in predicting breast cancer risk. In 1989, Hsu et al. reported no increase in the number of chromatid breaks per cell in 82 breast cancer cases compared to 335 controls (0.64 +/- 0.36 versus 0.60 +/-0.35, respectively) (4). We designed this current protocol with important differences from the earlier Hsu et al. study.

First, by evaluating only patients with a personal history of bilateral breast cancer (2/3 of whom also had a positive family history of breast cancer) we selected cases that had a greater probability of having a predisposing genotype. In the original Hsu study, patients with a history of a single breast cancer were selected without regard to age at diagnosis or family history status. A second important difference between the two studies was our use of γ-radiation as a mutagen compared to the bleomycin that was used in the earlier study.

The rationale for reinvestigating the mutagen sensitivity assay in our breast cancer study population is as follows. While the majority of breast cancers are believed to develop independently of an individual's genotype, it is clear that family history of breast cancer, particularly in a premenopausal first-degree relative, is an important risk factor for the development of this disease. This increased risk is likely due to inheritance of a predisposing genotype. The specific genes contributing to this predisposition are unknown in most women with breast cancer and a positive family history. Less than 7% of all breast cancers are thought to occur in the setting of a germline mutation in BRCA1 or BRCA2 (2,3). A phenotype assay, such as the one described in this report, is not dependent on the discovery of these unknown genetic conditions. An assay that can capture a common downstream functional effect of a variety of predisposing mutations would be relevant to a much broader population of women than a genotype sequencing approach.

A possible shortcoming of using radiation-induced chromatid-breaks as a predictor for breast cancer development is that this phenotype may not be a consistent consequence of all predisposing genetic conditions. For example, there is no evidence that individuals with Li-Fraumeni syndrome (a germline mutation in p53) have increased

susceptibility to chromatid-breaks. However, mutations in BRCA1, BRCA2, and ATM all affect cellular radiosensitivity and the success of double-strand break repair following ionizing radiation. Specifically, both BRCA1 and Brca2 colocalize with Rad51 following radiation-induced double-strand injuries (6,7). Additionally, normal function of BRCA1 is required for transcription-coupled repair following damage from ionizing radiation (8). Finally, BRCA1 has also been shown to associate with hRad50-hMre11-p95 in directing a cellular DNA damage response following ionizing radiation (9). A third tumor suppressor gene that may have relevance to breast cancer formation, ATM, also plays a critical role in the successful repair of DNA strand breaks following radiation (10). This role may in part be explained by the finding that BRCA1 protein function is dependent on phosphorylation by the ATM protein (11). It is clear that homozygous mutations in any of these three genes (BRCA1, BRCA2, or ATM) result in a radiosensitive phenotype (7-12).

The second rationale for using radiation as a mutagen for our experiment is that ionizing radiation is the most clearly recognized environmental carcinogen for breast cancer. The first evidence of the carcinogenic effect of radiation came from longitudinal studies of Japanese atomic bomb survivors (13). The importance of radiation as a breast carcinogen was further confirmed by the findings of increased breast cancer rates in women treated with radiation for nonmalignant conditions such as tuberculosis and enlargement of the thymus (14,15). The use of radiation as a cancer treatment also has been shown to carry carcinogenic risks. In a large study of girls treated with mantle irradiation for Hodgkin's disease, the 30-year actuarial risk of developing breast cancer

approached 35% (16). Furthermore, there appeared to be a dose-response relationship between radiation exposure and breast cancer development.

Our finding that radiation-induced chromatid breaks correlated with the risk of having bilateral breast cancer is in part supported by animal studies. Ponnaiya et al. noted that the significantly higher rates in radiation-induced mammary carcinoma in BALB/c mice compared to C57BL/6 mice correlated with differences in radiation-induced genomic instability in mammary epithelial tissue (17). After 16 population doublings, irradiated mammary cells from BALB/c mice had significantly more chromatid breaks than C57BL/6 mice. These data suggested that a genotype that increases breast cancer susceptibility correlated with a phenotype of sensitivity to radiation-induced chromatid breaks.

Following Hsu et al.'s initial investigation of chromatid breaks in breast cancer patients, a number of other investigators have also evaluated whether a chromatid break assay could predict the risk of breast cancer development. Three series with relatively small number of breast cancer cases all showed an increase in the median number of induced lymphocyte chromatid breaks in cases versus controls (18-20). In the largest series to date, Scott et al. found a statistically significant increase in chromatid breaks per cell in 135 women with a single breast cancer compared to 105 controls with no breast cancer history (21). Together these data, along with our current study, suggest the phenotype of sensitivity to radiation-induced chromatid breaks correlates with the risk of breast cancer development. However, as shown in Figure 1, there is a considerable degree of overlap in the assay results between cases and controls. This suggests that the assay is unlikely to develop into a test with high-sensitivity and high-specificity. None-

the-less, the test may be of clinical value for an individual found to have a high number of chromatid breaks. In our study, a value of 0.65 or greater captured 40% of the cases compared with only 5% of the controls.

Two studies that have investigated radiation-induced chromatid breaks in first-degree relatives have provided further evidence that the radiosensitivity noted in breast cancer cases is genetically based. Patel et al. reported that first-degree relatives of breast cancer patients had more radiation-induced chromosome breaks compared to controls (20). Additionally, Roberts et al. recently reported that 62% of first-degree relatives of 16 radiosensitive breast cancer patients from the Scott et al. study (21) were also radiosensitive (22). This compared to a rate of only 7% in first-degree relatives of 4 breast cancer patients with a low number of chromatid breaks per cell (22). Furthermore, Roberts et al. modeled the inheritance pattern of radiosensitivity and breast cancer and suggested that the data fit with a marker of an inherited low-penetrance breast cancer predisposition gene(s).

A potential shortcoming of the lymphocyte assay that we used in this study is that lymphocyte response to radiation is likely dependent on a number of factors. For example, it is possible that cytokines, released either from cancer cells or in response to having cancer, can affect lymphocyte response. To more precisely distinguish the genetic and epigenetic influences on lymphocyte chromatid breaks, more data comparing the rates of the mutagen sensitivity assay in individuals with known predisposing genotypes, individuals with single breast cancers and no family history and individuals without a cancer history will be needed.

In conclusion, increasing data suggest that screening for the phenotype of radiation-induced chromatid breaks may prove useful as a biological predictor for breast cancer risk. We believe that the preliminary data in this report needs additional confirmatory data, as the aggregate data of our study and those reported in the literature is relatively small and is subject to publication bias (negative studies of this type are unlikely to be reported).

References

1. Greenlee RT, Murray T, Bolden S, et al.: Cancer Statistics, 2000. *CA Cancer J Clin* 2000;50:7-33.

2. Newman B, Mu H, Butler LM, et al.: Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. *JAMA* 1998;279:915-921.

3. Ford D, Easton DF. The genetics of breast and ovarian cancer. *Br J Cancer* 1995;72:805-812.

4. Hsu TC, Johnston DA, Cherry LA, et al.: Sensitivity to genotoxic effects of bleomycin in humans: possible relationship to environmental carcinogenesis. *Int J Cancer* 1989;43:403-409.

5. Shattuck-Eidens D, Oliphant A, McClure M, et al.: BRCA1 sequence analysis in women at high risk for susceptibility mutations. *JAMA* 1997;278:1242-1250.

6. Scully R, Chen J, Plug A, et al.: Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell* 1997;88:265-275.

7. Sharan SK, Morimatsu M, Albrecht U, et al.: Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature* 1997;386:804-810.

8. Gowen LC, Avrutskaya AV, Latour AM, et al. BRCA1 required for transcription-coupled repair of oxidative DNA damage. *Science* 1998;281:1009-1012.

9. Zhong Q, Chen CF, Li S, et al. Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. *Science* 1999;285(5428):747-750.

10. Blocher D, Sigut D, Hannan MA: Fibroblasts for ataxia telangiectasia (AT) and AT heterozygotes show an enhanced level of residual DNA double-strand breaks after low dose-rate γ-irradiation as assayed by pulsed field gel electrophoresis. *Int J Radiat Biol* 1991;60:791-802.

11. Cortez D, Wang Y, Qin J, et al.: Requirement of ATM-dependent phosphorylation of Brca1 in the DNA damage response to double-strand breaks. *Science* 1999;5(286):1162-1166.

12. Morimatsu M, Donoho G, Hasty P: Cells deleted for Brca2 COOH terminus exhibit hypersensitivity to γ-radiation and premature senescence. *Cancer Res* 1998;58:3441-3447.

13. Tokunaga M, Land CE, Tokuoka S, et al.: Incidence of female breast cancer among atomic bomb survivors, 1950-1985. *Rad Res* 1994;138(2):209-223.

14. Hildreth NG, Shore RE, Dvoretsky PM: The risk of breast cancer after irradiation of the thymus in infancy. *N Engl J Med* 1989:321(19):1281-1284.

15. Hrubec Z, Boice JD, Monson RR, et al.: Breast cancer after multiple chest fluoroscopies: second follow-up of Massachusetts women with tuberculosis. *Cancer Res* 1989;49(1):229-234.

16. Bhatia S, Robison LL, Oberlin O, et al.: Breast cancer and other second neoplasms after childhood Hodgkin's disease. *New Eng J Med* 1996;334(12):745-51.

17. Ponnaiya B, Cornforth MN, Ullrich RL: Radiation-induced chromosomal instability in BALB/c and C57BL/6 mice: the difference is as clear as black and white. *Rad Res* 1997;147:121-125.

18. Parshad R, Price FM, Bohr VA, et al: Deficient DNA repair capacity, a predisposing factor in breast cancer. *Br J Cancer* 1996;74(1):1-5.

19. Helzlsouer KJ, Harris EL, Parshad R, et al.: DNA repair proficiency: potential susceptibility factor for breast cancer. *J Natl Ca Inst* 1996;88 (11):754-755.

20. Patel RK, Trived AH, Arora DC, et al: DNA repair proficiency in breast cancer patients and their first-degree relatives. *Int J Cancer* 1997;73:20-24.

21. Scott D, Barber JBP, Spreadborough AR, et al.: Increased chromosomal radiosensitivity in breast cancer patients: a comparison of two assays. *Int J Radiat Biol* 1999;75 (1):1-10.

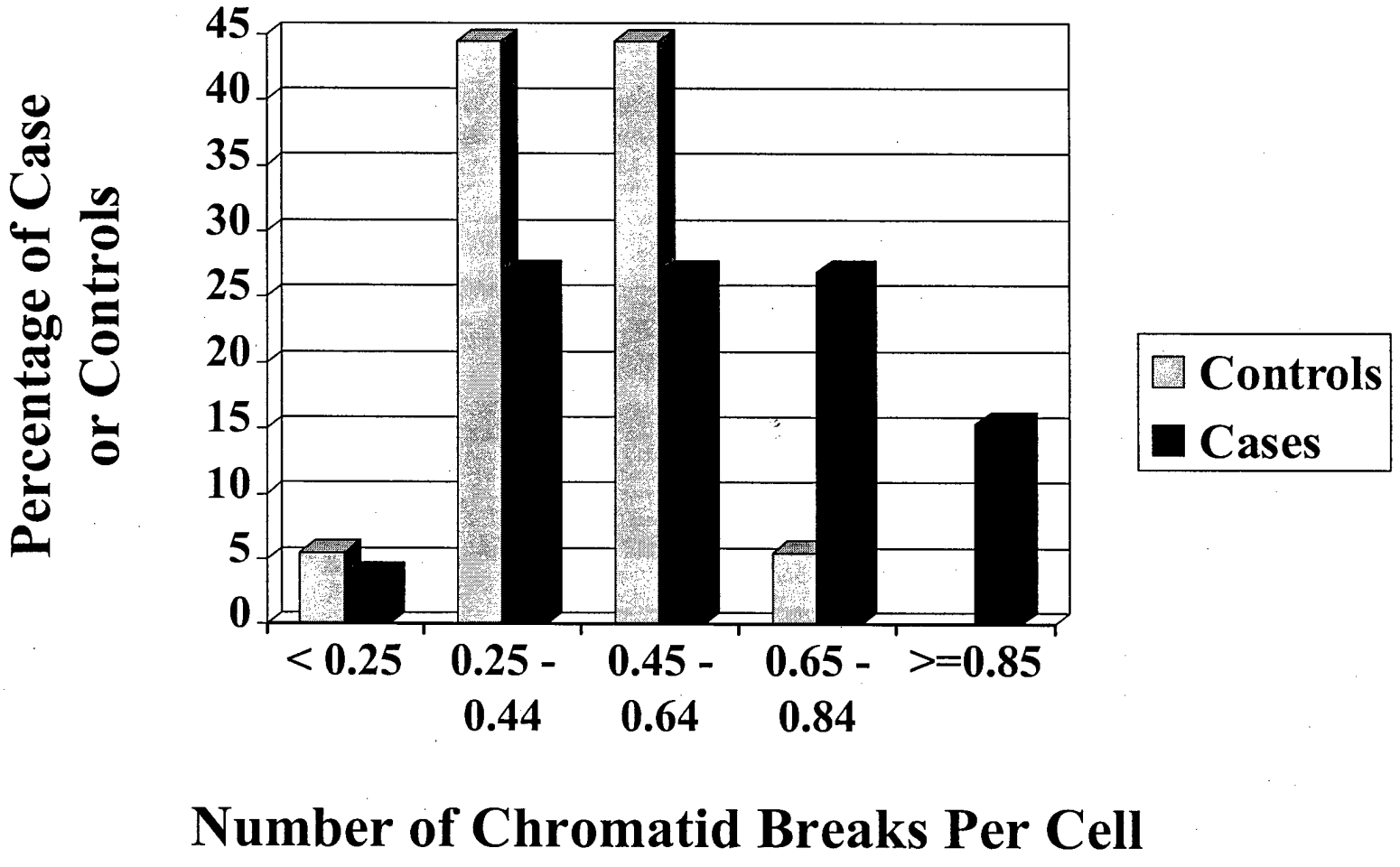
22. Roberts SA, Spreadborough AR, Bulman B, et al.: Heritability of cellular radiosensitivity: a marker of low-penetrance predisposition genes in breast cancer? *Am J Hum Genet* 1999;65:784-794.

Figure Legends

Figure 1, Distribution of the case and control populations as a function of the mutagen sensitivity assay results.

Figure 2, Distribution of the bilateral breast cancer patients with a positive (+ FH) or negative (- FH) breast cancer family history as a function of the mutagen sensitivity assay results.

Cases Versus Controls



Cases: Family History